



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

sPlotOpen – An environmentally balanced, open-access, global dataset of vegetation plots

Sabatini, Francesco Maria ; Lenoir, Jonathan ; Hattab, Tarek ; Arnst, Elise Aimee ; Chytrý, Milan ;
Dengler, Jürgen ; et al ; Kessler, Michael

Abstract: Motivation: Assessing biodiversity status and trends in plant communities is critical for understanding, quantifying and predicting the effects of global change on ecosystems. Vegetation plots record the occurrence or abundance of all plant species co-occurring within delimited local areas. This allows species absences to be inferred, information seldom provided by existing global plant datasets. Although many vegetation plots have been recorded, most are not available to the global research community. A recent initiative, called ‘sPlot’, compiled the first global vegetation plot database, and continues to grow and curate it. The sPlot database, however, is extremely unbalanced spatially and environmentally, and is not open-access. Here, we address both these issues by (a) resampling the vegetation plots using several environmental variables as sampling strata and (b) securing permission from data holders of 105 local-to-regional datasets to openly release data. We thus present sPlotOpen, the largest open-access dataset of vegetation plots ever released. sPlotOpen can be used to explore global diversity at the plant community level, as ground truth data in remote sensing applications, or as a baseline for biodiversity monitoring. Main types of variable contained: Vegetation plots ($n = 95,104$) recording cover or abundance of naturally co-occurring vascular plant species within delimited areas. sPlotOpen contains three partially overlapping resampled datasets (c. 50,000 plots each), to be used as replicates in global analyses. Besides geographical location, date, plot size, biome, elevation, slope, aspect, vegetation type, naturalness, coverage of various vegetation layers, and source dataset, plot-level data also include community-weighted means and variances of 18 plant functional traits from the TRY Plant Trait Database. Spatial location and grain: Global, 0.01–40,000 m². Time period and grain: 1888–2015, recording dates. Major taxa and level of measurement: 42,677 vascular plant taxa, plot-level records. Software format: Three main matrices (.csv), relationally linked.

DOI: <https://doi.org/10.1111/geb.13346>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-204558>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Sabatini, Francesco Maria; Lenoir, Jonathan; Hattab, Tarek; Arnst, Elise Aimee; Chytrý, Milan; Dengler, Jürgen; et al; Kessler, Michael (2021). sPlotOpen – An environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecology and Biogeography*, 30(9):1740-1764.
DOI: <https://doi.org/10.1111/geb.13346>

sPlotOpen – An environmentally balanced, open-access, global dataset of vegetation plots

Francesco Maria Sabatini^{1,2}  | Jonathan Lenoir³  | Tarek Hattab⁴  |
 Elise Aimee Arnst⁵  | Milan Chytrý⁶  | Jürgen Dengler^{1,7,8}  | Patrice De Ruffray⁹ |
 Stephan M. Hennekens¹⁰  | Ute Jandt^{1,2}  | Florian Jansen¹¹  |
 Borja Jiménez-Alfaro¹²  | Jens Kattge¹³  | Aurora Levesley¹⁴ | Valério D. Pillar¹⁵  |
 Oliver Purschke¹⁶  | Brody Sandel¹⁷  | Fahmida Sultana¹⁸ | Tsipe Aavik¹⁹  |
 Svetlana Aćić²⁰  | Alicia T. R. Acosta²¹  | Emiliano Agrillo²²  | Miguel Alvarez²³  |
 Iva Apostolova²⁴ | Mohammed A. S. Arfin Khan²⁵  | Luzmila Arroyo²⁶ | Fabio Attorre²⁷  |
 Isabelle Aubin²⁸  | Arindam Banerjee²⁹ | Marijn Bauters^{30,31}  |
 Yves Bergeron³²  | Erwin Bergmeier³³  | Idoia Biurrun³⁴  | Anne D. Bjorkman^{35,36}  |
 Gianmaria Bonari³⁷  | Viktoria Bondareva³⁸  | Jörg Brunet³⁹  | Andraž Čarni^{40,41}  |
 Laura Casella⁴²  | Luis Cayuela⁴³  | Tomáš Černý⁴⁴ | Victor Chepinoga⁴⁵  |
 János Csiky⁴⁶  | Renata Ćušterevska⁴⁷ | Els De Bie⁴⁸  | André Luis de Gasper⁴⁹  |
 Michele De Sanctis²⁷  | Panayotis Dimopoulos⁵⁰ | Jiri Dolezal⁵¹  | Tetiana Dziuba⁵² |
 Mohamed Abd El-Rouf Mousa El-Sheikh^{53,54}  | Brian Enquist⁵⁵  | Jörg Ewald⁵⁶  |
 Farideh Fazayeli^{57,58} | Richard Field⁵⁹  | Manfred Finckh⁶⁰  | Sophie Gachet⁶¹  |
 Antonio Galán-de-Mera^{62,63,64}  | Emmanuel Garbolino⁶⁵ | Hamid Gholizadeh⁶⁶  |
 Melisa Giorgis⁶⁷  | Valentin Golub⁶⁸  | Inger Greve Alsos⁶⁹  | John-Arvid Grytnes⁷⁰ |
 Gregory Richard Guerin⁷¹  | Alvaro G. Gutiérrez⁷²  | Sylvia Haider^{1,2}  |
 Mohamed Z. Hatim^{73,74}  | Bruno Hérault^{75,76,77}  | Guillermo Hinojos Mendoza⁷⁸ |
 Norbert Hölzel⁷⁹  | Jürgen Homeier⁸⁰  | Wannes Hubau^{81,82} | Adrian Indreica⁸³ |
 John A. M. Janssen⁸⁴ | Birgit Jedrzejek⁷⁹ | Anke Jentsch⁸⁵  | Norbert Jürgens⁶⁰  |
 Zygmunt Kącki⁸⁶ | Jutta Kapfer⁸⁷ | Dirk Nikolaus Karger⁸⁸  | Ali Kavgacı⁸⁹  |
 Elizabeth Kearsley⁹⁰  | Michael Kessler⁹¹  | Larisa Khanina⁹²  | Timothy Killeen⁹³ |
 Andrey Korolyuk⁹⁴ | Holger Kreft⁹⁵  | Hjalmar S. Kühl^{1,96}  | Anna Kuzemko⁹⁷  |
 Flavia Landucci⁶  | Attila Lengyel⁹⁸  | Frederic Lens^{99,100}  |
 Débora Vanessa Lingner¹⁰¹  | Hongyan Liu¹⁰²  | Tatiana Lysenko^{103,104,105}  |
 Miguel D. Mahecha^{1,106}  | Corrado Marcenò^{6,34}  | Vasiliy Martynenko¹⁰⁷  |
 Jesper Erenskjold Moeslund¹⁰⁸  | Abel Monteagudo Mendoza^{109,110}  |

Francesco Maria Sabatini and Jonathan Lenoir contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Global Ecology and Biogeography* published by John Wiley & Sons Ltd.

Ladislav Mucina^{111,112}  | Jonas V. Müller¹¹³  | Jérôme Munzinger¹¹⁴  |
 Alireza Naqinezhad¹¹⁵ | Jalil Noroozi¹¹⁶  | Arkadiusz Nowak^{117,118}  |
 Viktor Onyshchenko¹¹⁹ | Gerhard E. Overbeck¹²⁰  | Meelis Pärtel¹⁹  |
 Aníbal Pauchard^{121,122}  | Robert K. Peet¹²³  | Josep Peñuelas^{124,125}  |
 Aaron Pérez-Haase^{126,127}  | Tomáš Peterka⁶ | Petr Petřík¹²⁸  | Gwendolyn Peyre¹²⁹  |
 Oliver L. Phillips¹⁴  | Vadim Prokhorov¹³⁰ | Valerijus Rašomavičius¹³¹  |
 Rasmus Revmann^{132,133}  | Gonzalo Rivas-Torres¹³⁴  | John S. Rodwell¹³⁵ |
 Eszter Ruprecht¹³⁶ | Solvita Rūsiņa¹³⁷  | Cyrus Samimi¹³⁸ | Marco Schmidt¹³⁹  |
 Franziska Schrodtt⁵⁹  | Hanhuai Shan¹⁴⁰ | Pavel Shirokikh¹⁰⁷  | Jozef Šibík¹⁴¹  |
 Urban Šilc¹⁴²  | Petr Sklenář¹⁴³ | Željko Škvorc¹⁴⁴  | Ben Sparrow¹⁴⁵ |
 Marta Gaia Sperandii^{21,146}  | Zvezdana Stančić¹⁴⁷ | Jens-Christian Svenning¹⁴⁸  |
 Zhiyao Tang¹⁰² | Cindy Q. Tang¹⁴⁹ | Ioannis Tsiripidis¹⁵⁰ | Kim André Vanselow¹⁵¹  |
 Rodolfo Vásquez Martínez¹⁰⁹ | Kiril Vassilev²⁴ | Eduardo Vélez-Martin¹⁵²  |
 Roberto Venanzoni¹⁵³  | Alexander Christian Vibrans¹⁰¹ | Cyrille Violle¹⁵⁴  |
 Risto Virtanen^{1,155,156}  | Henrik von Wehrden¹⁵⁷ | Viktoria Wagner¹⁵⁸  |
 Donald A. Walker¹⁵⁹  | Donald M. Waller¹⁶⁰  | Hua-Feng Wang¹⁶¹ |
 Karsten Wesche^{1,162,163} | Timothy J. S. Whitfeld¹⁶⁴  | Wolfgang Willner¹¹⁶  |
 Susan K. Wiser⁵  | Thomas Wohlgemuth¹⁶⁵  | Sergey Yamalov¹⁶⁶ |
 Martin Zobel¹⁹  | Helge Bruelheide^{1,2} 

¹German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

²Institute of Biology, Martin-Luther University Halle-Wittenberg, Halle, Germany

³UMR CNRS 7058 "Ecologie et Dynamique des Systèmes Anthropisés" (EDYSAN), Université de Picardie Jules Verne, Amiens, France

⁴MARBEC, Univ Montpellier, CNRS, IFREMER and IRD, Sète, France

⁵Manaaki Whenua - Landcare Research, Lincoln, New Zealand

⁶Department of Botany and Zoology, Masaryk University, Brno, Czech Republic

⁷Vegetation Ecology Group, Institute of Natural Resource Sciences (IUNR), Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland

⁸Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany

⁹Institut de biologie moléculaire des plantes-CNRS, Université de Strasbourg, Strasbourg, France

¹⁰Wageningen Environmental Research, Wageningen, the Netherlands

¹¹Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany

¹²Research Unit of Biodiversity (CSIC/UO/PA), University of Oviedo, Mieres, Spain

¹³Max Planck Institute for Biogeochemistry, Jena, Germany

¹⁴School of Geography, University of Leeds, Leeds, UK

¹⁵Department of Ecology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

¹⁶Institute for Medical Epidemiology, Biometrics and Informatics (IMEBI), Interdisciplinary Center for Health Sciences, Medical School of the Martin-Luther University Halle-Wittenberg, Halle/Saale, Germany

¹⁷Department of Biology, Santa Clara University, Santa Clara, California, USA

¹⁸Shahjalal University of Science & Technology, Sylhet, Bangladesh

¹⁹Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia

²⁰Faculty of Agriculture, Department of Botany, University of Belgrade, Belgrade-Zemun, Serbia

²¹Department of Sciences, Roma Tre University, Rome, Italy

²²Biodiversity Conservation Department, ISPRA - Italian National Institute for Environmental Protection and Research, Rome, Italy

²³University of Bonn, INRES, Bonn, Germany

²⁴Department of Plant and Fungal Diversity and Resources, Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

²⁵Shahjalal University of Science & Technology, Sylhet, Bangladesh

- ²⁶Dirección de la Carrera de Biología, Universidad Autónoma Gabriel René Moreno, Santa Cruz de la Sierra, Bolivia
- ²⁷Department of Environmental Biology, Sapienza University of Rome, Rome, Italy
- ²⁸Natural Resources Canada, Great Lakes Forestry Centre, Canadian Forest Service, Sault Ste Marie, Ontario, Canada
- ²⁹Department of Computer Science, University of Illinois Urbana Champaign, Urbana, Illinois, USA
- ³⁰Department Green Chemistry and Technology, Isotope Bioscience Laboratory (UGent-ISOFYS), Ghent University, Ghent, Belgium
- ³¹Department Environment, Computational and Applied Vegetation Ecology (UGent-CAVELab), Ghent University, Ghent, Belgium
- ³²Forest Research Institute, Université du Québec en Abitibi-Témiscamingue, Rouyn-Noranda, Quebec, Canada
- ³³Vegetation Ecology and Phytodiversity, University of Göttingen, Göttingen, Germany
- ³⁴Plant Biology and Ecology, University of the Basque Country UPV/EHU, Bilbao, Spain
- ³⁵Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden
- ³⁶Gothenburg Global Biodiversity Centre, Gothenburg, Sweden
- ³⁷Free University of Bozen-Bolzano, Bolzano, Italy
- ³⁸Laboratory of Phytodiversity Problem and of Phytocoenology, Institute of Ecology of the Volga River Basin, Toljatty, Russian Federation
- ³⁹Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences, Alnarp, Sweden
- ⁴⁰Institute of Biology, Research Center of the Slovenian Academy of Sciences and Arts, Ljubljana, Slovenia
- ⁴¹School for Viticulture and Enology, University of Nova Gorica, Nova Gorica, Slovenia
- ⁴²Biodiversity Conservation Department, ISPRA - Italian National Institute for Environmental Protection and Research, Rome, Italy
- ⁴³Department of Biology and Geology, Physics and Inorganic Chemistry, Universidad Rey Juan Carlos, Móstoles, Spain
- ⁴⁴Department of Forest Ecology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Praha 6 - Suchbát, Czech Republic
- ⁴⁵Central Siberian Botanical Garden SB RAS, Novosibirsk, Russian Federation
- ⁴⁶Department of Ecology, University of Pécs, Pécs, Hungary
- ⁴⁷Institute of Biology, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Skopje, Republic of North Macedonia
- ⁴⁸Biotope Diversity, Research Institute for Nature and Forest (INBO), Brussels, Belgium
- ⁴⁹Universidade Regional de Blumenau, Blumenau, Brazil
- ⁵⁰Laboratory of Botany, Division of Plant Biology, Department of Biology, University of Patras, Patras, Greece
- ⁵¹Department of Functional Ecology, Institute of Botany, Czech Academy of Sciences, Trebon, Czech Republic
- ⁵²M.G. Kholodny Institute of Botany, National Academy of Sciences of Ukraine, Geobotany and Ecology, Kyiv, Ukraine
- ⁵³Botany and Microbiology Department, College of Science, King Saud University, Riyadh, Saudi Arabia
- ⁵⁴Botany Department, Faculty of Science, Damanshour University, Damanshour, Egypt
- ⁵⁵Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA
- ⁵⁶Hochschule Weihenstephan-Triesdorf, University of Applied Sciences, Freising, Germany
- ⁵⁷Google LLC, Mountain View, California, USA
- ⁵⁸University of Minnesota - Twin Cities, Minneapolis, Minnesota, USA
- ⁵⁹School of Geography, University of Nottingham, Nottingham, UK
- ⁶⁰Biodiversity, Ecology and Evolution of Plants, Institute for Plant Science & Microbiology, University of Hamburg, Hamburg, Germany
- ⁶¹CNRS, IRD, IMBE, Aix Marseille Univ, Avignon Université, Marseille, France
- ⁶²Laboratorio de Botánica, Universidad CEU San Pablo, Madrid, Spain
- ⁶³Laboratorio de Botánica, Universidad Privada Antonio Guillermo Urrelo, Cajamarca, Peru
- ⁶⁴Herbario AQP, Estudios Fitogeográficos del Perú, Paucarpata, Arequipa, Peru
- ⁶⁵Nova Sophia - Regus Nova, Climact Data Science (CDS), CS, Sophia Antipolis Cedex, France
- ⁶⁶Department of Plant Biology, University of Mazandaran, Babolsar, Iran
- ⁶⁷Ecología Vegetal y Fitogeografía, Instituto Multidisciplinario de Biología Vegetal (IMBIV-CONICET), Córdoba, Argentina
- ⁶⁸Laboratory of Phytocoenology, Samara Federal Research Center of the Russian Academy of Sciences, Institute of Ecology of the Volga river basin of the Russian Academy of Science, Toljatty, Russian Federation
- ⁶⁹The Arctic University Museum of Norway, UiT - The Arctic University of Norway, Tromsø, Norway
- ⁷⁰Department of Biological Sciences, University of Bergen, Bergen, Norway
- ⁷¹School of Biological Sciences, University of Adelaide, Adelaide, Australia
- ⁷²Departamento de Ciencias Ambientales y Recursos Naturales Renovables, Facultad de Ciencias Agronomicas, Universidad de Chile, Santiago, Chile
- ⁷³Plant Ecology and Nature Conservation Group - Environmental Sciences Department, Wageningen University, Wageningen, the Netherlands
- ⁷⁴Botany and Microbiology Department - Faculty of Science, Tanta University, Tanta, Egypt
- ⁷⁵CIRAD, UPR Forêts et Sociétés, Yamoussoukro, Ivory Coast
- ⁷⁶University of Montpellier, CIRAD, Montpellier, France
- ⁷⁷INP-HB, Institut National Polytechnique Félix Houphouët-Boigny, Yamoussoukro, Ivory Coast

- ⁷⁸ASES Ecological and Sustainable Services, Aubenas, France
- ⁷⁹Institute of Landscape Ecology, University of Münster, Münster, Germany
- ⁸⁰Plant Ecology and Ecosystems Research, University of Göttingen, Göttingen, Germany
- ⁸¹Department Environment, Laboratory of Wood Biology (UGent-WoodLab), Ghent University, Ghent, Belgium
- ⁸²Service of Wood Biology, Royal Museum for Central Africa, Tervuren, Belgium
- ⁸³Department of Silviculture, Transilvania University of Brasov, Brasov, Romania
- ⁸⁴Wageningen University and Research, Wageningen Environmental Research (Alterra), Wageningen, the Netherlands
- ⁸⁵Disturbance Ecology, Bayreuth Center of Ecology and Environmental Research, University of Bayreuth, Bayreuth, Germany
- ⁸⁶Botanical Garden, University of Wrocław, Wrocław, Poland
- ⁸⁷Norwegian Institute of Bioeconomy Research, Tromsø, Norway
- ⁸⁸Biodiversity and Conservation Biology, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
- ⁸⁹Faculty of Forestry, Kilavuzlar Köyü Öte Karsi Üniversite Kampüsü Merkez, Karabuk University, Karabuk, Turkey
- ⁹⁰Department Environment, Computational and Applied Vegetation Ecology (UGent-CAVELab), Ghent University, Ghent, Belgium
- ⁹¹Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland
- ⁹²Branch of the M.V. Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Institute of Mathematical Problems of Biology of RAS, Pushchino, Russian Federation
- ⁹³Museo de Historia Natural Noel Kempff Mercado, Universidad Autonoma Gabriel Rene Moreno, Santa Cruz de la Sierra, Bolivia
- ⁹⁴Geosystem Laboratory, Central Siberian Botanical Garden, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russian Federation
- ⁹⁵Department of Biodiversity, Macroecology and Biogeography, University of Göttingen, Göttingen, Germany
- ⁹⁶Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
- ⁹⁷Department of Geobotany and Ecology, M.G. Kholodny Institute of Botany of the National Academy of Sciences of Ukraine, Kyiv, Ukraine
- ⁹⁸Centre for Ecological Research, Institute of Ecology and Botany, Vácrátót, Hungary
- ⁹⁹Research Group Functional Traits, Naturalis Biodiversity Center, Leiden, the Netherlands
- ¹⁰⁰Institute of Biology Leiden, Leiden University, Leiden, the Netherlands
- ¹⁰¹Departamento de Engenharia Florestal, Universidade Regional de Blumenau, Blumenau, Brazil
- ¹⁰²College of Urban and Environmental Sciences, Peking University, Beijing, China
- ¹⁰³Laboratory of Vegetation Science, Komarov Botanical Institute RAS, Saint-Petersburg, Russian Federation
- ¹⁰⁴Laboratory of Phytodiversity Problems, Institute of Ecology of the Volga River Basin RAS - Branch of the Samara Scientific Center RAS, Togliatti, Russian Federation
- ¹⁰⁵Group of Ecology of Living Organisms, Tobolsk complex scientific station of Ural Branch RAS, Tobolsk, Russian Federation
- ¹⁰⁶Remote Sensing Centre for Earth System Research, University of Leipzig, Leipzig, Germany
- ¹⁰⁷Ufa Institute of Biology, Ufa Federal Scientific Center of the Russian Academy of Sciences, Ufa, Russian Federation
- ¹⁰⁸Department of Bioscience, Aarhus University, Roende, Denmark
- ¹⁰⁹Jardín Botánico de Missouri Oxapampa, Oxapampa, Pasco, Peru
- ¹¹⁰Universidad Nacional de San Antonio Abad del Cusco, Cusco, Peru
- ¹¹¹Murdoch University, Murdoch, Perth, Western Australia, Australia
- ¹¹²Department of Geography & Environmental Studies, Stellenbosch University, Stellenbosch, South Africa
- ¹¹³Conservation Science, Royal Botanic Gardens, Kew, Ardingly, UK
- ¹¹⁴CIRAD, CNRS, INRAE, AMAP, Université de Montpellier, Montpellier, France
- ¹¹⁵Department of Plant Biology, University of Mazandaran, Mazandaran, Iran
- ¹¹⁶Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria
- ¹¹⁷Botanical Garden - Center for Biodiversity Conservation, Polish Academy of Sciences, Warsaw, Poland
- ¹¹⁸Institute of Biology, University of Opole, Opole, Poland
- ¹¹⁹National Academy of Sciences of Ukraine, M.G. Kholodny Institute of Botany, Kyiv, Ukraine
- ¹²⁰Department of Botany, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
- ¹²¹Laboratorio de Invasiones Biológicas (LIB), Facultad de Ciencias Forestales, Universidad de Concepción, Concepción, Chile
- ¹²²Institute of Ecology and Biodiversity (IEB), Santiago, Chile
- ¹²³Department of Biology, University of North Carolina, Chapel Hill, North Carolina, USA
- ¹²⁴Global Ecology Unit CSIC-CREAF-UAB, CSIC, Bellaterra, Catalonia, Spain
- ¹²⁵CREAF, Cerdanyola del Valles, Spain
- ¹²⁶Department of Biosciences, University of Vic-Central University of Catalonia, Barcelona, Spain
- ¹²⁷Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Barcelona, Spain
- ¹²⁸Department of Vegetation Ecology, Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic

- ¹²⁹Department of Civil and Environmental Engineering, University of the Andes, Bogota, Colombia
- ¹³⁰Institute of Environmental Sciences, Kazan Federal University, Kazan, Russian Federation
- ¹³¹Institute of Botany, Nature Research Centre, Vilnius, Lithuania
- ¹³²Biodiversity, Ecology and Evolution of Plants/Institute for Plant Science & Microbiology, University of Hamburg, Hamburg, Germany
- ¹³³Faculty of Natural Resources and Spatial Sciences, Namibia University of Science and Technology, Windhoek, Namibia
- ¹³⁴Universidad San Francisco de Quito, COCIBA, Quito, Ecuador
- ¹³⁵Independent Researcher, Lancaster, UK
- ¹³⁶Hungarian Department of Biology and Ecology, Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania
- ¹³⁷Faculty of Geography and Earth Sciences, University of Latvia, Riga, Latvia
- ¹³⁸Climatology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany
- ¹³⁹Palmengarten, Stadt Frankfurt am Main - Der Magistrat, Frankfurt am Main, Germany
- ¹⁴⁰Microsoft, Redmond, Washington, USA
- ¹⁴¹Institute of Botany, Plant Science and Biodiversity Centre Slovak Academy of Sciences, Bratislava, Slovakia
- ¹⁴²Institute of Biology, Research Centre of Slovenian Academy of Sciences and Arts (ZRC SAZU), Ljubljana, Slovenia
- ¹⁴³Department of Botany, Charles University, Prague, Czech Republic
- ¹⁴⁴Faculty of Forestry and Wood Technology, University of Zagreb, Zagreb, Croatia
- ¹⁴⁵TERN, University of Adelaide, Adelaide, South Australia, Australia
- ¹⁴⁶CSIC-UV-GV, Centro de Investigaciones sobre Desertificación, Moncada, Spain
- ¹⁴⁷Faculty of Geotechnical Engineering, University of Zagreb, Varaždin, Croatia
- ¹⁴⁸Department of Biology, Aarhus University, Aarhus C, Denmark
- ¹⁴⁹School of Ecology and Environmental Science, Yunnan University, Chenggong New District, Kunming, China
- ¹⁵⁰School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece
- ¹⁵¹Department of Geography, University of Erlangen-Nuremberg, Erlangen, Germany
- ¹⁵²ILEX Consultoria Científica, Porto Alegre, Brazil
- ¹⁵³Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy
- ¹⁵⁴Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, CEFE, Montpellier, France
- ¹⁵⁵Ecology and Genetics Research Unit, Biodiversity Unit, University of Oulu, Oulu, Finland
- ¹⁵⁶Department of Physiological Diversity, Helmholtz Center for Environmental Research - UFZ, Leipzig, Germany
- ¹⁵⁷Institute of Ecology, Leuphana University of Lüneburg, Lüneburg, Germany
- ¹⁵⁸Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada
- ¹⁵⁹Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska, USA
- ¹⁶⁰Botany, University of Wisconsin-Madison, Madison, Wisconsin, USA
- ¹⁶¹Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou, China
- ¹⁶²Botany Department, Senckenberg Museum of Natural History Görlitz, Görlitz, Germany
- ¹⁶³International Institute Zittau, Technische Universität Dresden, Zittau, Germany
- ¹⁶⁴Bell Museum, University of Minnesota, St. Paul, Minnesota, USA
- ¹⁶⁵Forest Dynamics, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
- ¹⁶⁶Laboratory of Wild-Growing Flora, South-Ural Botanical Garden-Institute, Ufa Scientific Centre, Russian Academy of Sciences, Ufa, Russian Federation

Correspondence

Francesco Maria Sabatini, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, 04103, Leipzig, Germany.
Email: francesco.sabatini@botanik.uni-halle.de

Funding information

Agence Nationale de la Recherche, Grant/Award Number: ANR-07-BDIV-0006, ANR-07-BDIV-0008 and ANR-07-BDIV-0010; H2020 European Research Council, Grant/Award Number: ERC Advanced Grant 291585 "T-FORCES" and ERC-SyG-2013-610028 IMBALANCE-P; Villum Fonden, Grant/Award Number: 16549; Deutsche Forschungsgemeinschaft, Grant/Award Number: DFG FZT 118, 202548816, DFG Ho3296-2, DFG Ho3296-4, DFG VA 749/1-1, DFG VA 749/4-1 and DFG WE 2601/3-1 3-2 4-1 4-2; Narodowe Centrum Nauki, Grant/Award Number: 2017/25/B/NZ8/00572; Latvia grant, Grant/Award Number: AAP2016/B041//Zd2016/AZ03; NSF, Grant/Award Number: DEB-0415383; Horizon 2020 Framework Programme, Grant/Award Number: 640176; U.S. National Science Foundation, Grant/Award Number: DBI-0213794 and DBI-9905838; Grantová Agentura České Republiky, Grant/Award Number: 19-28491X; German Centre for Integrative Biodiversity Research, Grant/Award Number: 50170649_#7; Fundación BBVA, Grant/Award Number: BIOCON08_044; Akademie Věd České Republiky, Grant/Award Number: RVO 67985939; Spanish Research Agency, Grant/Award Number: AEI/10.13039/501100011033; National Research, Development and Innovation Office, Hungary, Grant/Award Number: PD-12399; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 20BD21_184131; Basque Government, Grant/Award Number: 640176; Russian Foundation for Basic Research, Grant/Award Number: 16-04-00747a; Brazil's National Council of Scientific and Technological Development, Grant/Award Number: 307689/2014-0 and 310022/2015-0; Volkswagen Foundation, Grant/Award Number: AZ I/81 976

Handling Editor: Amanda Bates

Abstract

Motivation: Assessing biodiversity status and trends in plant communities is critical for understanding, quantifying and predicting the effects of global change on ecosystems. Vegetation plots record the occurrence or abundance of all plant species co-occurring within delimited local areas. This allows species absences to be inferred, information seldom provided by existing global plant datasets. Although many vegetation plots have been recorded, most are not available to the global research community. A recent initiative, called 'sPlot', compiled the first global vegetation plot database, and continues to grow and curate it. The sPlot database, however, is extremely unbalanced spatially and environmentally, and is not open-access. Here, we address both these issues by (a) resampling the vegetation plots using several environmental variables as sampling strata and (b) securing permission from data holders of 105 local-to-regional datasets to openly release data. We thus present sPlotOpen, the largest open-access dataset of vegetation plots ever released. sPlotOpen can be used to explore global diversity at the plant community level, as ground truth data in remote sensing applications, or as a baseline for biodiversity monitoring.

Main types of variable contained: Vegetation plots ($n = 95,104$) recording cover or abundance of naturally co-occurring vascular plant species within delimited areas. sPlotOpen contains three partially overlapping resampled datasets (c. 50,000 plots each), to be used as replicates in global analyses. Besides geographical location, date, plot size, biome, elevation, slope, aspect, vegetation type, naturalness, coverage of various vegetation layers, and source dataset, plot-level data also include community-weighted means and variances of 18 plant functional traits from the TRY Plant Trait Database.

Spatial location and grain: Global, 0.01–40,000 m².

Time period and grain: 1888–2015, recording dates.

Major taxa and level of measurement: 42,677 vascular plant taxa, plot-level records.

Software format: Three main matrices (.csv), relationally linked.

KEYWORDS

big data, biodiversity, biogeography, database, functional traits, macroecology, vascular plants, vegetation plots

1 | BACKGROUND & SUMMARY

Biodiversity is facing a global crisis. As many as 1 million species are currently threatened with extinction, the vast majority due to anthropogenic impacts such as land-use and climate change (IPBES, 2019; WWF, 2020). In addition, the rates of biodiversity homogenization and redistribution are accelerating (Fricke & Svenning, 2020; Lenoir et al., 2020; Staude et al., 2020). Biological assemblages are becoming progressively more similar to each other globally, as local and endemic species go extinct and are replaced by more widespread and competitive native or alien species (IPBES, 2019; Staude et al., 2020). Many terrestrial and marine species are also shifting their geographical distribution as a response to climate change

(Lenoir et al., 2020). This has profound potential impacts on ecosystems and human health (Bonebrake et al., 2018; Pecl et al., 2017).

Plant communities are no exception to this biodiversity crisis (Cardinale et al., 2011; Lenoir et al., 2008; Staude et al., 2020). This is particularly worrying since terrestrial vegetation accounts for 80% (450 Gt C) of the living biomass on Earth (Bar-On et al., 2018). Given the central role of vegetation in ecosystem productivity, structure, stability and functioning (Cardinale et al., 2011), assessing biodiversity status and trends in plant communities is paramount for other kingdoms of life and human societies alike.

Monitoring trends in plant biodiversity requires adequate data across a range of spatio-temporal scales (Kühl et al., 2020; Pimm, 2021). Large independent collections of plant occurrence

data do exist at the global or continental extent via the Botanical Information and Ecology Network (BIEN; Enquist et al., 2016), the Global Inventory of Floras and Traits (GIFT; Weigelt et al., 2020) or the Global Biodiversity Information Facility (GBIF; <https://www.gbif.org/>). However, these databases suffer from one or several of the following limitations: (a) imbalance towards tree species only; (b) lack of data on how individual plant species co-occur and interact locally to form plant communities; and (c) coarse spatial resolutions (e.g., one-degree grid cells), which preclude intersection with high resolution remote sensing data and the assessment of biodiversity trends at the plant community level (Boakes et al., 2010).

There is a long tradition among botanists and phytosociologists to record the cover or abundance of each plant species that occurs in a vegetation plot (here used as a synonym of 'relevé' or 'quadrat') of a given size (i.e., surface area) at a given time (e.g., Stebler & Schröter, 1892). Compared to presence-only data, vegetation-plot data present many advantages. As all visible plant species are recorded, plots contain information on which plant species do, and do not co-occur in the same locality at a given moment in time (Chytrý et al., 2016). This is important for testing hypotheses related to biotic interactions among plant species. Vegetation-plot data also provide crucial information on where and when a species was absent, therefore, improving predictions from current species distribution models (Phillips et al., 2009). Being spatially explicit, vegetation plots can be resurveyed through time to assess potential changes in plant species composition relative to a baseline (Perring et al., 2018; Staude et al., 2020; Steinbauer et al., 2018). As they normally contain information on the relative cover or abundance of each species, vegetation plots are also more appropriate for detecting biodiversity changes than data representing only the occurrence of individual species (Beck et al., 2018; Jandt et al., 2011).

Globally, however, vegetation-plot data are very fragmented, as they typically stem from a myriad of local research and survey projects (Bruehlheide et al., 2019). These are fine-grained data (e.g., 1–10,000 m²) normally covering small spatial extents (e.g., 1–1,000 km²). With their disparate sampling protocols, standards and taxonomic resolutions, aggregating and harmonizing vegetation plot data proves extremely challenging (Bruehlheide et al., 2018). It is not surprising, therefore, that these data are rarely used in global-scale research on the biodiversity of plant communities (Aubin et al., 2020; Franklin et al., 2017; Wiser, 2016).

The sPlot initiative tries to close this data gap. It consolidates numerous local to regional vegetation-plot datasets to create a harmonized and comprehensive global database of georeferenced terrestrial plant species assemblages (Bruehlheide et al., 2019). Established in 2013, sPlot v3.0 currently contains more than 1.9 million vegetation plots, and is fully integrated with the TRY database (Kattge et al., 2020), from which it derives information on plant functional traits. The sPlot database is increasingly being used to study continental-to-global scale vegetation patterns (Cai et al., 2021; Testolin, Attorre, et al., 2021; Testolin, Carmona, et al., 2021), such as the relative contribution of regional versus local factors to the global patterns of fern richness (Weigand et al., 2020), the

mechanisms underlying the spread and abundance of native versus invasive tree species (van der Sande et al., 2020), and worldwide trait–environment relationships in plant communities (Bruehlheide et al., 2018).

Yet, most of these data are not open-access. Here, we secured permission from data holders in the sPlot database to openly release a dataset composed of 95,104 vegetation plots. We selected the plots to be released using a replicated environmental stratification, in order to represent the entire environmental space covered by the sPlot database. This maximizes the benefits of releasing these data for a wide range of potential uses. The selected vegetation plots stem from 105 databases and span 114 countries (Figure 1). This resampled dataset (sPlotOpen – hereafter) is composed of: (a) plot-level information, including metadata and basic vegetation structure descriptors; (b) the vascular plant species composition of each vegetation plot, including species cover or abundance information when available; and (c) community-level functional information obtained by intersection with the TRY database (Kattge et al., 2020).

sPlotOpen is specifically designed for global macroecological studies, for example, the exploration of functional diversity patterns of communities with continental-to-global extent. We expect, however, that sPlotOpen might likewise prove useful to answer a range of different questions, related for instance to species co-occurrence patterns, the definition of species pools, the link between regional versus local determinants of species diversity, or the niche overlap between co-occurring species. Yet, data in sPlotOpen should not be considered as representative of the distribution of plant communities worldwide, especially when working at local spatial extents. This should be kept in mind for applications such as species distribution models (SDMs) or joint SDMs, whose results might be affected by the uneven geographical distribution of sPlotOpen's data. We refer the reader to the section 'Usage notes' for additional guidance on critical issues related, for instance, to incompletely sampled vegetation plots, varying plot size, and nested vegetation plots.

2 | METHODS

2.1 | Vegetation plot data sources

We started from the sPlot database v2.1 (created in October 2016), which contains 1,121,244 unique vegetation plots and 23,586,216 species records. Most of the data in sPlot refer to natural and semi-natural vegetation, while vegetation shaped by intensive and repeated human interference, such as cropland or ruderal communities, is hardly represented. Data originate from 110 different vegetation-plot datasets of regional, national or continental extent, some of which stem from regional or continental initiatives (see Bruehlheide et al., 2019, for more information). For instance: 48 vegetation-plot datasets derive from the European Vegetation Archive (EVA; Chytrý et al., 2016); three major African datasets derive from the Tropical African Vegetation Archive (TAVA); and multiple vegetation datasets in the USA and Australia derive from the VegBank (Peet,

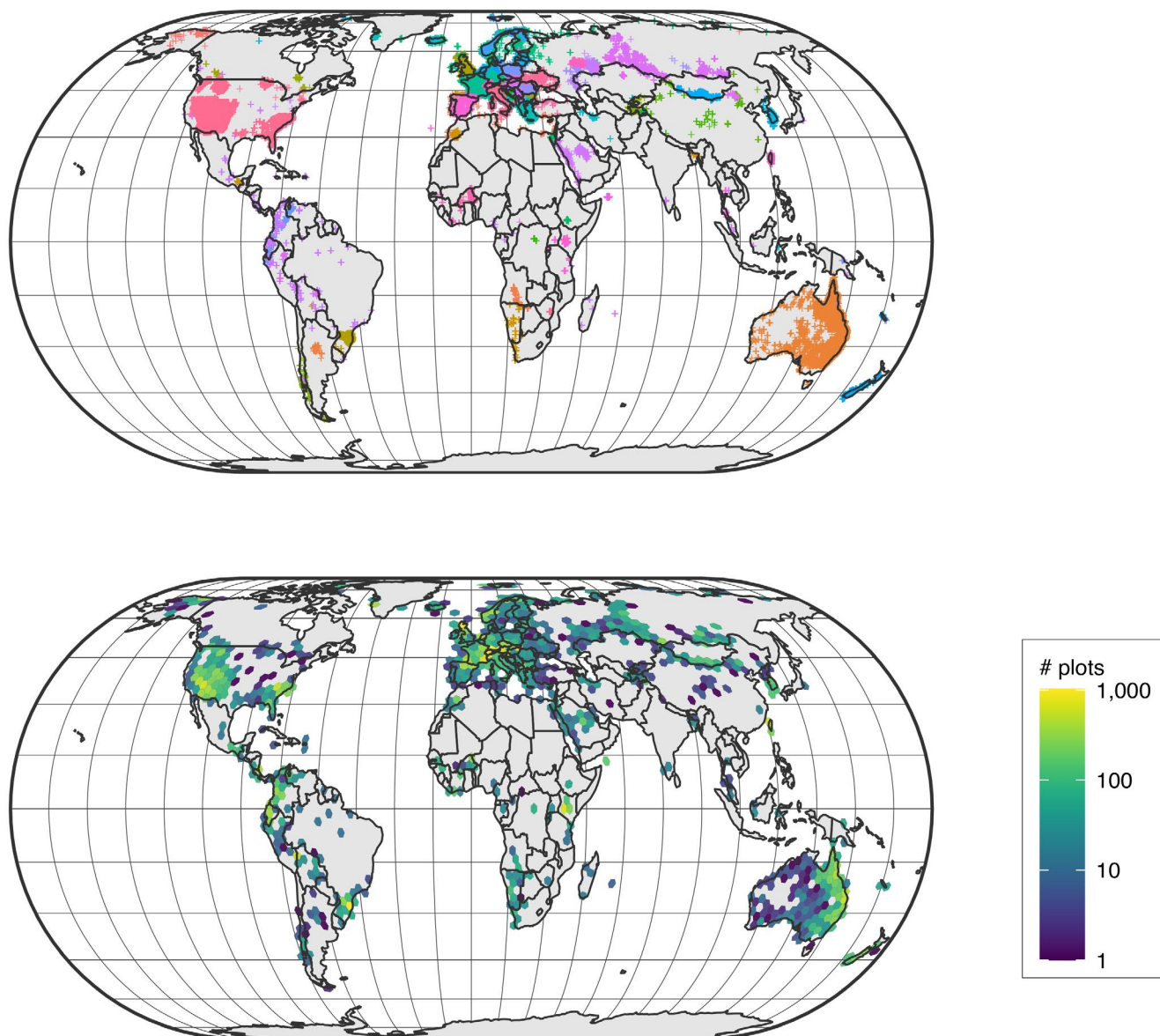


FIGURE 1 Top: global distribution of all vegetation plots contained in sPlotOpen ($n = 95,104$). Each colour represents a different source dataset ($n = 105$ – different datasets might have the same colour). Bottom: spatial distribution of vegetation plot density for the environmentally balanced dataset selected by the first resampling iteration ($n = 49,787$). Densities are calculated in hexagonal cells with a spatial resolution of approximately $70,000 \text{ km}^2$. Map projection is Eckert IV

Lee, Boyle, et al., 2012; Peet, Lee, Jennings, et al., 2012) and TERN's AEKOS (Chabbi & Loescher, 2017) archives, respectively. Data from other continents (South America, Asia) or countries were contributed as separate standalone datasets. The metadata of each individual vegetation-plot dataset stored in sPlot are managed through the Global Index of Vegetation-Plot Databases (GIVD; Dengler et al., 2011), using the GIVD code as the unique dataset identifier.

2.2 | Resampling method

Data in the sPlot database are unevenly distributed across vegetation types and geographical regions (see Bruelheide et al., 2018).

Mid-latitude regions in developed countries (mostly Europe, the USA and Australia) are overrepresented in sPlot, while regions in the tropics and subtropics are underrepresented, which is a typical geographical bias in biodiversity data (see Lenoir et al., 2020; Lenoir & Svenning, 2015 for similar geographical bias in species redistribution). Such a geographical bias usually translates into an environmental bias with temperate climate usually more represented than tropical or Mediterranean climates. Unbalanced sampling effort in the environmental space is of particular concern for comparative macroecological studies (Bruelheide et al., 2018; Lenoir et al., 2010). To reduce this imbalance as much as possible, we performed a stratified resampling approach within the environmental space using several environmental variables available at global extent as sampling strata.

First, we removed vegetation plots without geographical coordinates or with a location uncertainty higher than 3 km. We also removed vegetation plots identified by the respective data contributors as having been recorded in wetlands or in anthropogenic vegetation types, since these data were available only for a few geographical regions, mostly in Europe. This resulted in a total of 799,400 out of the initial set of 1,121,244 vegetation plots.

We then ran a global principal component analysis (PCA) on a matrix of all terrestrial grid cells at a spatial resolution of 2.5 arcmin ($n = 8,384,404$), based on 30 climatic and soil variables. For climate, we used the 19 bioclimatic variables from CHELSA (Climatologies at high resolution for the earth's land surface areas) v1.2 (Karger et al., 2017), as well as two other bioclimatic variables reflecting the growing-season length (growing degree days above 1°C – GDD1 – and 5°C – GDD5), which were derived from CHELSA's monthly temperatures as in Synes and Osborne (2011). In addition, we considered an index of aridity and a layer for potential evapotranspiration from the Consortium of Spatial Information (CGIAR-CSI, Trabucco & Zomer, 2010). For soil, we extracted seven variables from the SoilGrids database (Hengl et al., 2017), namely: (a) soil organic carbon

content in the fine earth fraction; (b) cation exchange capacity; (c) pH; as well as the fractions of (d) coarse fragments; (e) sand; (f) silt; and (g) clay. The results of this PCA represent the full environmental space of all terrestrial habitats on Earth, irrespective of whether a grid cell hosted vegetation plots or not (Supporting Information Figure S1). We then subdivided the PCA ordination space, represented by the first two principal components (PC1–PC2), which accounted for 47 and 23%, respectively, of the total environmental variation in terrestrial grid cells, into a regular 100×100 grid. This PC1–PC2 two-dimensional space was subsequently used to balance our sampling effort across all PC1–PC2 grid cells for which vegetation plots were available. After excluding 42,878 vegetation plots for which no PC1 or PC2 values were available, due to missing data in the bioclimatic or soil variables, we projected the remaining 756,522 vegetation plots onto this PC1–PC2 grid. We finally calculated how many vegetation plots occurred in each PC1–PC2 grid cell (Figure 2).

In total, vegetation plots were available for 1,720 out of the 4,125 PC1–PC2 grid cells covered by the 8,384,404 terrestrial grid cells of the geographical space. We then resampled those PC1–PC2 grid cells ($n = 858$) with more than 50 vegetation plots, which is

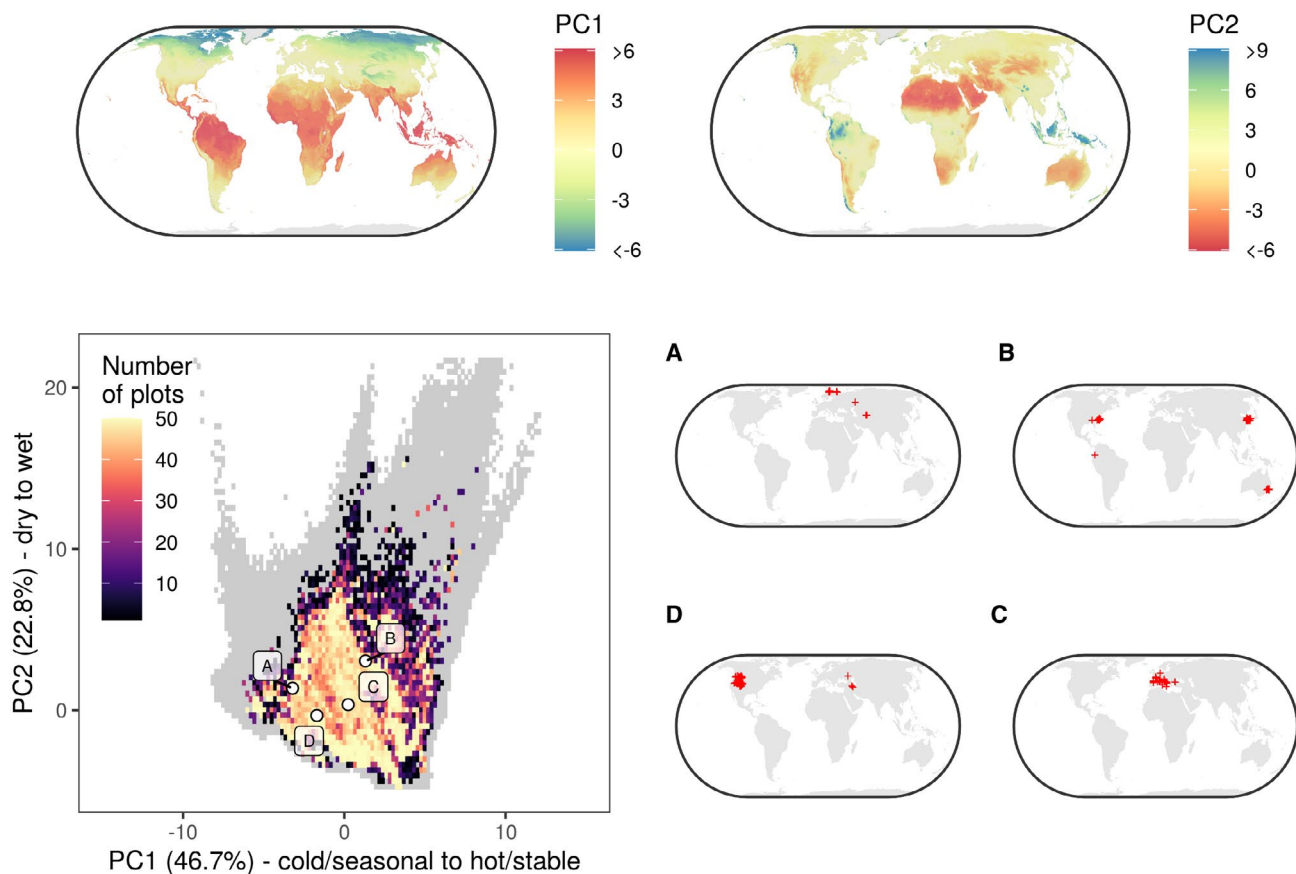


FIGURE 2 Distribution of vegetation plots from sPlotOpen in the global environmental space based on a principal component analysis (PCA) using 30 climate and soil variables. Top: spatial distribution of PCA values across all terrestrial grid cells ($n = 8,384,404$, spatial grain = 2.5 arcmin). Bottom left: distribution of plots compared to the distribution of all terrestrial 2.5 arc-minute cells (grey background) in the PCA space. Only the plots in the environmentally balanced dataset selected in the first resampling iteration are shown ($n = 49,787$). The PCA space was divided into a 100×100 regular grid. The first and second PCA axes explained 47 and 23% of the total variance, respectively. Bottom right: geographical distribution of the vegetation plots contained in four randomly selected PCA grid cells

the median number of plots occurring across occupied grid cells in sPlot. This threshold of 50 vegetation plots represents a compromise between selecting a high number of plots, and keeping the resampled dataset as balanced as possible across the PC1–PC2 environmental space. To select these 50 vegetation plots we used the heterogeneity-constrained random resampling algorithm (Lengyel et al., 2011). This algorithm quantifies the variability in plant species composition among a set of vegetation plots by computing the mean and the variance of the Jaccard's dissimilarity index (Jaccard, 1912) between all possible pairs of vegetation plots. More precisely, for a given PC1–PC2 grid cell containing more than 50 vegetation plots, we generated 1,000 random selections of 50 vegetation plots and ranked each selection according to the mean (ascending order) and variance (descending order) value of the Jaccard's dissimilarity index. Ranks from both sortings were summed for each random selection, and the selection with the lowest summed rank was considered to provide the most balanced/even representation of vegetation types within the focal grid cell. Where a grid cell contained fewer than 50 plots, we retained all of them. In this way, we reduced the imbalance towards over-sampled climate types while ensuring that the resampled dataset represents the entire environmental gradient covered by the original sPlot database. This approach optimizes the selection of a subset of vegetation plots that encompasses the highest variability in species composition while avoiding peculiar and rare communities, which may represent outliers. As such, our approach maximizes variability over representativeness within each grid cell. We repeated the whole resampling procedure three times to get three different environmentally balanced, resampled subsets of our vegetation plots. These three resampling iterations can therefore be used as separate replicates, albeit these are not completely independent, as the same plots might have been drawn in two or even three of the three resampling iterations. In addition, those plots located in PC1–PC2 grid cells with fewer than 50 vegetation plots are completely shared by all three iterations.

2.3 | Permission to release the data as open access

The resampling procedure resulted in 56,486, 56,501 and 56,494 vegetation plots selected during resampling iterations #1, #2 and #3, respectively, for a total of 107,238 unique vegetation plots. Since the sPlot database is a consortium of independent datasets whose copyright belongs to the data contributors, we used this preliminary potential selection to ask each dataset's custodian (i.e., either the owner of a dataset or its authorized representative in the case of a collective dataset) for permission to release the data of selected vegetation plots as open access. For 12,134 unique vegetation plots, permission could not be granted because, for instance, the data are unpublished, confidential or sensitive. The number of vegetation plots for which the open-access permission was not granted in resampling iterations #1, #2 and #3 was 6,699, 6,690 and 6,705, respectively.

To mitigate the imbalance due to the exclusion of these confidential plots, we created a 'consensus' dataset. We started from resampling iteration #1, and replaced the 6,699 plots not granted as open access with plots selected in the second and third iterations, for which such permission could be granted ('reserve' plots, hereafter). We imposed the constraint that each candidate vegetation plot in the reserve pool should belong to the same environmental stratum, that is, the same PC1–PC2 grid cell, as the confidential vegetation plot, even though we acknowledge that this procedure does not maximize the variability in plant species composition of the replacement plots. Even after drawing from reserves, there were 3,150 plots that could not be replaced. These were distributed across 279 PC1–PC2 grid cells (16.2% of occupied cells), each cell having on average 11 irreplaceable plots (min. = 1, median = 5, max. = 50).

2.4 | Trait information

For each vegetation plot for which open access could be granted, we computed the community-weighted mean and variance for 18 plant functional traits derived from the TRY database v3.0 (Kattge et al., 2020). These traits were selected among those that describe the leaf, wood, and seed economics spectra (Reich, 2014; Westoby, 1998), and are known to either affect different key ecosystem processes or respond to macroclimatic drivers, or both (Bruehlheide et al., 2018). The 18 plant functional traits (all concentrations based on dry weight) were: (a) leaf area (mm^2); (b) stem specific density (g/cm^3); (c) specific leaf area (m^2/kg); (d) leaf carbon concentration (mg/g); (e) leaf nitrogen concentration (mg/g); (f) leaf phosphorus concentration (mg/g); (g) plant height (m); (h) seed mass (mg); (i) seed length (mm); (j) leaf dry matter content (g/g); (k) leaf nitrogen per area (g/m^2); (l) leaf N:P ratio (g/g); (m) leaf $\delta^{15}\text{N}$ (per million); (n) seed number per reproductive unit; (o) leaf fresh mass (g); (p) stem conduit density (per mm^2); (q) dispersal unit length (mm); and (r) conduit element length (μm).

Because missing values were particularly widespread in the species-trait matrix, we calculated community-weighted means using the gap-filled version of these traits we received from TRY (Kattge et al., 2020). Gap-filling was performed at the level of individual observations and relies on hierarchical Bayesian modelling (R package 'BHPMF' – Fazayeli et al., 2014; Schrodte et al., 2015) in R (R Core Team, 2020). This is a Bayesian machine learning approach, with no a priori assumptions, except for the data being missing completely at random. The algorithm 'learns' from the data, that is, if there was a phylogenetic signal in the data, this was used to fill the gaps but where no such signal was apparent, none was introduced. After gap-filling, we transformed to the natural logarithm all gap-filled trait values and averaged each trait by taxon (i.e., at species or genus level). The gap-filling approach was run only for species having at least one trait observation ($n = 21,854$). Additional information on the gap-filling procedure is available in Bruehlheide et al. (2019).

Community-weighted means (CWM) and variances (CWV) were calculated for every plant functional trait j and every vegetation plot k as follows (Enquist et al., 2015):

$$CWM_{j,k} = \sum_i^{n_k} p_{i,k} t_{ij} \quad (1)$$

$$CWV_{j,k} = \sum_i^{n_k} p_{i,k} (t_{ij} - CWM_{j,k})^2 \quad (2)$$

where n_k is the number of species with trait information in vegetation plot k , $p_{i,k}$ is the relative abundance of species i in vegetation plot k calculated as the species' fraction in cover or abundance of total cover or abundance, and t_{ij} is the mean value of species i for trait j .

3 | DATA RECORDS

sPlotOpen contains 95,104 unique vegetation plots from 105 constitutive datasets (Table 1) and from 114 countries covering all continents except Antarctica (Figure 1). This is the result of pooling together the three environmentally balanced datasets from resampling iterations #1, #2 and #3 containing 49,787, 49,811 and 49,789 plots, respectively, after excluding the set of plots for which open access could not be granted by data contributors. The number of plots shared across all three resampling iterations is 19,672, while 14,939 plots are shared between two iterations. Replacing confidential plots in resampling iteration #1 with reserves from the other two iterations in the same PC1–PC2 grid cell resulted in a consensus version containing 53,262 plots. sPlotOpen only contains the species composition of vascular plants; information on the composition of bryophytes and lichens was discarded since it was only available for a minority of plots ($n = 11,001$ and $n = 6,801$, respectively). Information on the size (surface area) of the vegetation survey is available for 67,022 plots, and ranges between 0.03 and 40,000 m² (mean = 377 m²; median = 100 m²). Specifically, sPlotOpen contains 12,894 plots with size smaller than 10 m², 25,742 with size 10–100 m², 24,750 plots with size 100–1,000 m² and 3,075 plots with size greater or equal to 1,000 m². Similarly, only for a minority of plots ($n = 24,167$) is information on the exact group of plants sampled in the field available (e.g., complete vegetation, only trees, only trees > 1 m height, and so on). However, as most data were collected using the phytosociological method, we deem it safe to assume that, unless otherwise specified, plots contain information on all vascular plants. We retained plots with incomplete vegetation, because they were mostly located in the tropics, that is, in areas where vegetation plots are particularly scarce otherwise. The average number of vascular plant species per vegetation plot ranges between 1 (i.e., mono-specific stands) and 271 species (mean = 20; median = 16).

By capping the number of vegetation plots in overrepresented environmental conditions, the resampling procedure described above strongly reduced the bias in the distribution of vegetation plots within the PC1–PC2 environmental space. Yet, due to the lack

or scarcity of data from some geographical regions, like the tropics, there is some remaining imbalance in the spatial distribution of vegetation plots across geographical regions (Figure 1). This is evident when comparing the number of plots across continents. When considering the first resampling iteration only ($n = 49,787$), Europe is by far the best represented continent, with 15,920 vegetation plots. The least represented continents are Africa and South America, with 3,709 and 5,498 vegetation plots, respectively. Some residual imbalance remains also when considering biomes (Figure 3). With the exception of the 'Temperate mid-latitudes' biome, which includes 14,100 vegetation plots, all other biomes have a number of plots comprised between 1,558 ('Polar and subpolar zone') and 6,245 ('Subtropics with year-round rain') vegetation plots (Figure 3, left). Despite this residual imbalance, all the Whittaker biomes are covered by sPlotOpen (Figure 3, right), and our resampling algorithm has resulted in a much more balanced dataset than many other global datasets that are available, such as GBIF.

Almost one third of the 95,104 vegetation plots in sPlotOpen belong to forests ($n = 38,282$), one half to non-forest vegetation ($n = 45,735$), with 11.6% of plots remaining unassigned ($n = 11,087$). When not directly done by data providers, the assignment of plots to forests and non-forests was based on multiple lines of evidence, including the plot-level information on the cover of the tree layer, as well as traits of species composing a plot, such as growth form and height. In short, a plot record was considered as forest if the cover of the tree layer, or alternatively, the sum of the (relative) cover of all tree taxa (scaled by the sum of all cover values, as a percentage), was greater than 25%. It was considered a non-forest record if the sum of relative cover of low-stature, non-tree and non-shrub taxa was greater than 90%. For an extensive explanation of this classification scheme, we refer the reader to Bruehlheide et al. (2019). Even though the proportion of forest versus non-forest vegetation plots is relatively well balanced, the geographical distribution of vegetation plots belonging to different vegetation types is likely not balanced in the geographical space, as it depends on the idiosyncrasies of the constitutive datasets composing the sPlot database. For instance, the data from New Zealand only include plots collected in non-forest ecosystems, while data from Chile only refer to forests. We urge potential users to carefully read the section 'Usage notes' below and the description of each individual dataset in GIVD (Dengler et al., 2011), and to contact the custodians of each dataset for further information.

4 | DATABASE ORGANIZATION

The environmentally balanced and open-access dataset sPlotOpen is organized into three main matrices, relationally linked through the key column 'PlotObservationID'.

The 'header' matrix contains plot-level information for the 95,104 vegetation plots, including: metadata (e.g., plot ID, data source, sampling date, geographical location, positional accuracy); sampling design information (e.g., the total surface area used during the vegetation survey); and a plot-level description of vegetation

TABLE 1 List of datasets contributing to sPlotOpen, the environmentally balanced and open-access database of vegetation plots

GIVD ID	Dataset name	Custodian	Deputy custodian	No. open-access plots	Reference
00-00-001	ForestPlots.net	Oliver L. Phillips	Aurora Levesley	169	Lopez-Gonzalez et al. (2011)
00-00-003	SALVIAS	Brian Enquist	Brad Boyle	3,403	
00-00-004	Vegetation Database of Eurasian Tundra	Risto Virtanen		519	
00-00-005	Tundra Vegetation Plots (TundraPlot)	Anne D. Bjorkman	Sarah Elmendorf	309	Elmendorf et al. (2012)
00-RU-001	Vegetation Database Forest of Southern Ural	Vasiliy Martynenko	Pavel Shirokikh	68	
00-RU-002	Database of Masaryk University's Vegetation Research in Siberia	Milan Chytrý		158	Chytrý (2012)
00-RU-003	Database Meadows and Steppes of Southern Ural	Sergey Yamalov	Mariya Lebedeva	238	
00-TR-001	Forest Vegetation Database of Turkey – FVDT	Ali Kavgacı		45	
AF-00-001	West African Vegetation Database	Marco Schmidt	Georg Zizka	258	Schmidt et al. (2012)
AF-00-003	BIOTA Southern Africa Biodiversity Observatories Vegetation Database	Norbert Jürgens	Ute Schmiedel	1,015	Muche et al. (2012)
AF-00-006	SWEA-Dataveg	Miguel Alvarez	Michael Curran	1,675	Alvarez et al. (2021)
AF-00-008	PANAF Vegetation Database	Hjalmar S. Kühl	TeneKwetché Sop	884	
AF-00-009	Vegetation Database of the Okavango Basin	Rasmus Revermann	Manfred Finckh	378	Revermann et al. (2016)
AF-BF-001	Sahel Vegetation Database	Jonas V. Müller	Marco Schmidt	556	Müller (2003)
AF-CD-001	Forest Database of Central Congo Basin	Kim Sarah Jacobsen	Hans Verbeeck	140	Kearsley et al. (2013)
AF-ET-001	Vegetation Database of Ethiopia	Desalegn Wana	Anke Jentsch	67	Wana & Beierkuhnlein (2011)
AF-MA-001	Vegetation Database of Southern Morocco	Manfred Finckh		621	Finckh (2012)
AF-ZW-001	Vegetation Database of Zimbabwe	Cyrus Samimi		31	Samimi (2003)
AS-00-001	Korean Forest Database	Tomáš Černý	Jiri Dolezal	1,039	Černý et al. (2015)
AS-00-003	Vegetation of Middle Asia	Arkadiusz Nowak	Marcin Nobis	314	Nowak et al. (2017)
AS-00-004	Rice Field Vegetation Database	Arkadiusz Nowak		32	
AS-BD-001	Tropical Forest Dataset of Bangladesh	Mohammed A. S. Arfin Khan	Fahmida Sultana	87	
AS-CN-001	China Forest-Steppe Ecotone Database	Hongyan Liu	Fengjun Zhao	117	Liu et al. (2000)
AS-CN-002	Tibet-PaDeMoS Grazing Transect	Karsten Wesche	Yun Jäschke	58	Wang et al. (2017)
AS-CN-003	Vegetation Database of the BEF China Project	Helge Bruehlheide		24	Bruehlheide et al. (2011)
AS-CN-004	Vegetation Database of the Northern Mountains in China	Zhiyao Tang		124	
AS-EG-001	Vegetation Database of Sinai in Egypt	Mohamed Z. Hatim		143	Hatim (2012)
AS-ID-001	Sulawesi Vegetation Database	Michael Kessler		24	
AS-IR-001	Vegetation Database of Iran	Jalil Noroozi	Parastoo Mahdavi	277	
AS-KZ-001	Database of Meadow Vegetation in the NW Tien Shan Mountains	Viktoria Wagner		13	Wagner (2009)
AS-MN-001	Southern Gobi Protected Areas Database	Henrik von Wehrden	Karsten Wesche	1,032	von Wehrden et al. (2009)
AS-RU-001	Wetland Vegetation Database of Baikal Siberia (WETBS)	Victor Chepinoga		9	Chepinoga (2012)

(Continues)

TABLE 1 (Continued)

GIVD ID	Dataset name	Custodian	Deputy custodian	No. open-access plots	Reference
AS-RU-002	Database of Siberian Vegetation (DSV)	Andrey Korolyuk	Andrei Zverev	3,634	Korolyuk & Zverev (2012)
AS-RU-004	Database of the University of Münster – Biodiversity and Ecosystem Research Group's Vegetation Research in Western Siberia and Kazakhstan	Norbert Hölzel	Wanja Mathar	207	
AS-SA-001	Vegetation Database of Saudi Arabia	Mohamed Abd El-Rouf Mousa El-Sheikh		711	El-Sheikh et al. (2017)
AS-TJ-001	Eastern Pamirs	Kim André Vanselow		221	Vanselow (2016)
AS-TW-001	National Vegetation Database of Taiwan	Ching-Feng Li	Chang-Fu Hsieh	912	
AS-YE-001	Socotra Vegetation Database	Michele De Sanctis	Fabio Attorre	236	De Sanctis & Attorre (2012)
AU-AU-002	AEKOS	Ben Sparrow		10,976	Chabbi & Loescher (2017)
AU-NC-001	New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN)	Jérôme Munzinger	Philippe Birnbaum	98	Ibanez et al. (2014)
AU-NZ-001	New Zealand National Vegetation Databank	Susan K. Wiser		1,127	Wiser et al. (2001)
AU-PG-001	Forest Plots from Papua New Guinea	Timothy J. S. Whitfeld	George D. Weiblen	60	Whitfeld et al. (2014)
EU-00-002	Nordic-Baltic Grassland Vegetation Database (NBGVD)	Jürgen Dengler	Łukasz Kozub	54	Dengler & Rüşniņa (2012)
EU-00-011	Vegetation-Plot Database of the University of the Basque Country (BIOVEG)	Idoia Biurrun	Itziar García-Mijangos	2,142	Biurrun et al. (2012)
EU-00-013	Balkan Dry Grasslands Database	Kiril Vassilev	Armin Macanović	269	Vassilev et al. (2012)
EU-00-016	Mediterranean Ammophiletea Database	Corrado Marcenò	Borja Jiménez-Alfaro	783	Marcenò & Jiménez-Alfaro (2017)
EU-00-017	European Coastal Vegetation Database	John A. M. Janssen		356	
EU-00-018	The Nordic Vegetation Database	Jonathan Lenoir	Jens-Christian Svenning	1,735	Lenoir et al. (2013)
EU-00-019	Balkan Vegetation Database	Kiril Vassilev	Hristo Pedashenko	484	Vassilev et al. (2016)
EU-00-020	WetVegEurope	Flavia Landucci		127	Landucci et al. (2015)
EU-00-022	European Mire Vegetation Database	Tomáš Peterka	Martin Jiroušek	2,560	Peterka et al. (2015)
EU-AL-001	Vegetation Database of Albania	Michele De Sanctis	Giuliano Fanelli	31	De Sanctis et al. (2017)
EU-AT-001	Austrian Vegetation Database	Wolfgang Willner	Christian Berg	2,310	Willner et al. (2012)
EU-BE-002	INBOVEG	Els De Bie		119	
EU-BG-001	Bulgarian Vegetation Database	Iva Apostolova	Desislava Sopotlieva	160	Apostolova et al. (2012)
EU-CH-005	Swiss Forest Vegetation Database	Thomas Wohlgemuth		2,134	Wohlgemuth (2012)
EU-CZ-001	Czech National Phytosociological Database	Milan Chytrý	Ilona Knollová	1,287	Chytrý & Rafajová (2003)
EU-DE-001	VegMV	Florian Jansen	Christian Berg	15	Jansen et al. (2012)
EU-DE-013	VegetWeb Germany	Florian Jansen	Jörg Ewald	587	Ewald et al. (2012)
EU-DE-014	German Vegetation Reference Database (GVRD)	Ute Jandt	Helge Bruehlheide	762	Jandt & Bruehlheide (2012)

(Continues)

TABLE 1 (Continued)

GIVD ID	Dataset name	Custodian	Deputy custodian	No. open-access plots	Reference
EU-DK-002	National Vegetation Database of Denmark	Jesper Erenskjold Moeslund	Rasmus Ejrnæs	332	
EU-ES-001	Iberian and Macaronesian Vegetation Information System (SIVIM) – Wetlands	Aaron Pérez-Haase	Xavier Font	580	
EU-FR-003	SOPHY	Emmanuel Garbolino	Patrice De Ruffray	7,986	Garbolino et al. (2012)
EU-GB-001	UK National Vegetation Classification Database	John S. Rodwell		3,182	
EU-GR-001	KRITI	Erwin Bergmeier		22	
EU-GR-005	Hellenic Natura 2000 Vegetation Database (HelNatVeg)	Panayotis Dimopoulos	Ioannis Tsiripidis	620	Dimopoulos & Tsiripidis (2012)
EU-GR-006	Hellenic Woodland Database	Ioannis Tsiripidis	Georgios Fotiadis	17	Fotiadis et al. (2012)
EU-HR-001	Phytosociological Database of Non-Forest Vegetation in Croatia	Zvezdana Stančić		193	Stančić (2012)
EU-HR-002	Croatian Vegetation Database	Željko Škvorc	Daniel Krstonošić	585	
EU-HU-003	CoenoDat Hungarian Phytosociological Database	János Csiky	Zoltán Botta-Dukát	46	Lájer et al. (2008)
EU-IT-001	VegItaly	Roberto Venzoni	Flavia Landucci	754	Landucci et al. (2012)
EU-IT-010	Vegetation database of Habitats in the Italian Alps – HabItAlp	Laura Casella	Pierangela Angelini	247	Casella et al. (2012)
EU-IT-011	Vegetation-Plot Database Sapienza University of Rome (VPD-Sapienza)	Emiliano Agrillo	Fabio Attorre	967	Agrillo et al. (2017)
EU-LT-001	Lithuanian Vegetation Database	Valerijus Rašomavičius	Domas Uogintas	81	
EU-LV-001	Semi-natural Grassland Vegetation Database of Latvia	Solvita Rūsiņa		369	Rūsiņa (2012)
EU-MK-001	Vegetation Database of the Republic of Macedonia	Renata Čuštrevska		28	
EU-NL-001	Dutch National Vegetation Database	Stephan M. Hennekens	Joop H. J. Schaminée	1,098	Schaminée et al. (2006)
EU-PL-001	Polish Vegetation Database	Zygmunt Kącki	Grzegorz Swacha	692	Kącki & Śliwiński (2012)
EU-RO-007	Romanian Forest Database	Adrian Indreica	Pavel Dan Turtureanu	166	Indreica et al. (2017)
EU-RO-008	Romanian Grassland Database	Eszter Ruprecht	Kiril Vassilev	82	Vassilev et al. (2018)
EU-RS-002	Vegetation Database Grassland Vegetation of Serbia	Svetlana Ačić	Zora Dajić Stevanović	217	Ačić et al. (2012)
EU-RU-002	Lower Volga Valley Phytosociological Database	Valentin Golub	Andrey Chuvashov	383	Golub et al. (2012)
EU-RU-003	Vegetation Database of the Volga and the Ural Rivers Basins	Tatiana Lysenko		174	Lysenko et al. (2012)
EU-RU-011	Vegetation Database of Tatarstan	Vadim Prokhorov	Maria Kozhevnikova	206	Prokhorov et al. (2017)
EU-SI-001	Vegetation Database of Slovenia	Urban Šilc	Filip Kūzmič	1,029	Šilc (2012)
EU-SK-001	Slovak Vegetation Database	Milan Valachovič	Jozef Šibík	2,394	Šibík (2012)
EU-UA-001	Ukrainian Grasslands Database	Anna Kuzemko	Yulia Vashenyak	301	Kuzemko (2012)
EU-UA-006	Vegetation Database of Ukraine and Adjacent Parts of Russia	Viktor Onyshchenko	Vitaliy Kolomiychuk	96	
NA-00-002	Tree Biodiversity Network (BIOTREE-NET)	Luis Cayuela		241	Cayuela et al. (2012)

(Continues)

TABLE 1 (Continued)

GIVD ID	Dataset name	Custodian	Deputy custodian	No. open-access plots	Reference
NA-CA-003	Database of Timberline Vegetation in NW North America	Viktoria Wagner	Toby Spribille	63	Wagner et al. (2014)
NA-CA-004	Understory of Sugar Maple Dominated Stands in Quebec and Ontario (Canada)	Isabelle Aubin		13	Aubin et al. (2007)
NA-CA-005	Boreal Forest of Canada	Philippe Marchand	Yves Bergeron	57	Harper et al. (2003)
NA-GL-001	Vegetation Database of Greenland	Birgit Jedrzejek	Fred J. A. Daniëls	441	Sieg et al. (2006)
NA-US-002	VegBank	Robert K. Peet	Michael T. Lee	14,965	Peet, Lee, Jennings, et al. (2012)
NA-US-006	Carolina Vegetation Survey Database	Robert K. Peet	Michael T. Lee	3,263	Peet, Lee, Boyle, et al. (2012)
NA-US-014	Alaska-Arctic Vegetation Archive	Donald A. Walker	Amy Breen	771	Walker et al. (2016)
SA-00-002	VegPáramo	Gwendolyn Peyre	Xavier Font	2,010	Peyre et al. (2015)
SA-AR-002	Vegetation Database of Central Argentina	Melisa Giorgis	Alicia T. R. Acosta	86	
SA-BO-003	Bolivia Forest Plots	Michael Kessler	Sebastian Herzog	44	
SA-BR-002	Forest Inventory, State of Santa Catarina, Brazil (IFFSC Project)	Alexander Christian Vibrans	André Luís de Gasper	1,561	Vibrans et al. (2020)
SA-BR-003	Grasslands of Rio Grande do Sul, Brazil	Eduardo Vélez-Martin	Valério D. Pillar	306	
SA-BR-004	Grassland Database of Campos Sulinos	Gerhard E. Overbeck	Valério D. Pillar	147	
SA-CL-002	SSAForests_Plots_db	Alvaro G. Gutiérrez		155	
SA-CL-003	Chilean Park Transects – Fondecyt 1040528	Aníbal Pauchard	Alicia Marticorena	44	Pauchard et al. (2013)
SA-EC-001	Ecuador Forest Plot Database	Jürgen Homeier		166	

Note: Datasets are ordered based on their ID in the Global Index of Vegetation Databases (GVID ID).

structure (e.g., vegetation type, percentage cover of each vegetation layer), vegetation type, and naturalness level (i.e., whether a plot belongs to the same formation that would occupy the site without human interference). Plots in Europe are also classified according to the European Nature Information System (EUNIS) habitat classification (column 'ESY'), based on the habitat classification expert system (ESY, Chytrý et al., 2020). For each vegetation plot, we further provide information on the dataset it originates from, based on the IDs used in GIVD (Dengler et al., 2011). We also report four binary fields describing whether a plot belongs to the three resampling iterations (columns 'Resample_1', 'Resample_2', 'Resample_3'), or to the first resampling iteration after the inclusion of replacement plots (column 'Resample_1_consensus'). A brief summary of all the 47 variables in the header matrix is provided in Table 2.

The 'DT' matrix contains data on the species composition of each plot. It is structured in a long format and contains 1,945,384 records from 42,680 vascular plant taxa, mostly resolved at the species level. For each record, we report both the taxon name as originally contributed by the data custodian (column 'Original_species'), and the taxon

name after taxonomic standardization (column 'Species'). For details on the taxonomic standardization, please see section 'Technical validation' below. For each species we also provided cover/abundance values. These follow different standards across the datasets constituting the sPlot database. We, therefore, provide both the cover/abundance value as reported in the original data (column 'Original_abundance'), together with the abundance scale that was originally used (column 'Abundance_scale'). This can take seven values: 'CoverPerc' = percentage cover; 'pa' = presence-absence; 'x_BA' = basal area (m²/ha, only for woody species); 'x_IC' = individual count, that is, number of individuals in plot; 'x_SC' = stem count, that is, number of stems in plot; 'x_IV' = importance value index; and 'x_PF' = presence frequency. The great majority of entries, however, use the percentage cover scale ($n = 1,709,000$). Finally, for each entry, we calculated a 'Relative_cover', that is, the cover/abundance of a given taxon divided by the total cover/abundance of all taxa in that vegetation plot.

The 'CWM_CWV' matrix contains the community-weighted means and variances calculated for each of the 18 functional traits mentioned above. It also contains three additional columns. The

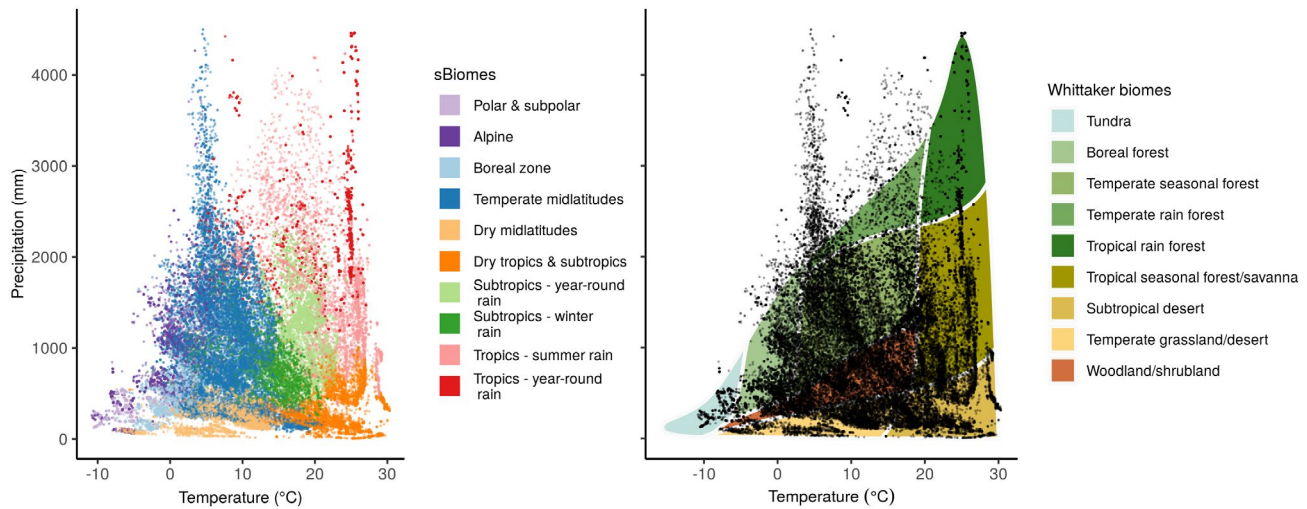


FIGURE 3 Distribution of vegetation plots in the first resampling iteration of sPlotOpen ($n = 49,787$) in the two-dimensional climatic space represented by mean annual temperature and mean annual precipitation. Left: plots are colour coded based on sBiomes, that is, sPlot's definition of biomes (Bruehlheide et al., 2019), which derives from Schultz's (2005) ecozones, modified to include also the alpine biome from Körner et al. (2017). Right: the same plots superimposed onto Whittaker's biomes (Whittaker, 1975), as adapted by Ricklefs (2008) and plotted using the R package 'plotbiomes'

column 'Species_richness' shows the number of species recorded in each plot. The columns 'Trait_coverage_cover' and 'Trait_coverage_pa' provide, respectively, the proportion of total cover and the proportion of species in a plot for which functional trait information was available. In total, functional trait information was available for 21,854 species. As functional trait information was based on gap-filled data (see above), each of these 21,854 species had information for all the 18 functional traits. The average proportion of species in each plot for which functional trait information was available is .85 (median = .95). For 42,012 plots, the coverage was complete, while we do not have functional trait information for any of the species occurring in 482 plots. When considering relative cover, the average trait coverage is .87, with 74,151 plots having functional trait information for species cumulatively accounting for more than 80% of relative cover. When considering the number of species, 68,041 plots have functional trait information for 80% or more of the species occurring in that plot.

sPlotOpen contains two additional objects. The 'metadata' matrix contains plot-level metadata, which provide information on the origin of each individual vegetation plot. This object contains 15 columns, with information on the dataset of origin (column 'GIVD_ID' – Dengler et al., 2011), author or surveyor names (columns 'Releve_author' and 'Releve_coauthor'), bibliographic references both at the dataset (column 'DB_BIBTEXKEY') and plot level ('Plot_Biblioreference' and 'BIBTEXKEY'), when available. Similarly, the column 'Project_name' provides information on the project in which a vegetation plot was originally recorded. When available, we also provide information on the numbering of the plots in the publication where they originally appeared (columns 'Nr_table_in_publ', 'Nr_releve_in_table'), or in the dataset where they were initially stored ('Original_nr_in_database'). In the case of nested plots ($n = 1,851$), we also provide the original plot and subplot IDs (columns: 'Original_plotID', 'Original_subplotID'). The last two columns

report plot-level 'Remarks', and the unique identifier produced by TURBOVEG when the vegetation plot was first stored ('GUID'). TURBOVEG is a program specifically designed to store, maintain and export vegetation plot data (<https://www.synbiosys.alterra.nl/turboveg>; Hennekens & Schaminée, 2001).

Finally, the object 'references' contains all the bibliographic references formatted according to a BibTex standard. Each reference is tagged with a key corresponding to the fields 'DB_BIBTEXKEY' and 'BIBTEXKEY' in the metadata. We further provide an R function ('sPlotOpen_citation') to create reference lists, based on a selection of plots and/or datasets.

Except for the 'reference' file (format.bib), all objects/matrices are provided in tab-delimited .txt files. All objects, including the 'sPlotOpen_citation' function, are also compiled inside a .RData object.

5 | TECHNICAL VALIDATION

The original sPlot database has a nested structure and consists of several individual datasets, each validated and maintained by its respective dataset custodian. In many cases, individual datasets are also collections whose vegetation plots were provided by their respective owners (the person who performed the actual vegetation survey) or by someone who digitized the original data from the scientific published or grey literature. We obviously have no direct control over the individual vegetation plots that we provide here in sPlotOpen. Yet, all these vegetation plots stem from trained professional botanists, or published scientific work, and are accompanied by detailed information on the sampling protocols used, thus ensuring data quality and reliability.

Before integration into the sPlot database, each dataset was further checked for consistency. If the dataset was in a different format, we converted it to a TURBOVEG 2 dataset (Hennekens & Schaminée,

TABLE 2 Description of the variables contained in the 'header' matrix, together with their range (if numeric) or possible levels (if nominal or binary) and the number of non-empty (i.e., non NA) records

Variable	Range/levels	Unit of measurement	No. of plots with information	Type
GIVD_ID	see Table 1		95,104	n
Dataset	see Table 1		95,104	n
Continent	Africa, Asia, Europe, North America, Oceania, South America		95,104	n
Country			95,104	n
Biome	Alpine, Boreal zone, Dry midlatitudes, Dry tropics and subtropics, Polar and subpolar zone, Subtropics with year-round rain, Subtropics with winter rain, Temperate midlatitudes, Tropics with summer rain, Tropics with year-round rain		95,104	n
Date_of_recording	05-07-1888 - 03-02-2015	dd-mm-yyyy	80,085	d
Latitude	-54.82303 - 80.149116	° (WGS84)	95,104	q
Longitude	-162.741433 - 176.4221	° (WGS84)	95,104	q
Location_uncertainty	1-2,750	m	95,075	q
Releve_area	0.03-40,000	m ²	67,022	q
Plant_recorded	All vascular plants, All trees & dominant understory, Dominant trees, Only dominant species, Dominant woody plants >= 2.5 cm dbh, All woody plants, Woody plants >= 1 cm dbh, Woody plants >= 2.5 cm dbh, Woody plants >= 5 cm dbh, Woody plants >= 10 cm dbh, Woody plants >= 20 cm dbh, Woody plants >= 1 m height, Not specified		95,104	n
Elevation	-30 - 5,960	m a.s.l.	62,968	q
Aspect	1-360	°	42,178	q
Slope	0-90	°	51,246	q
is_forest	FALSE = 45,735; TRUE = 38,282		84,017	b
ESY			39,632	n
Naturalness	1 = Natural, 2 = Semi-natural		60,192	o
Forest	FALSE = 36,282; TRUE = 33,170		69,452	b
Shrubland	FALSE = 58,245; TRUE = 11,207		69,452	b
Grassland	FALSE = 33,800; TRUE = 35,652		69,452	b
Wetland	FALSE = 59,196; TRUE = 10,256		69,452	b
Sparse_vegetation	FALSE = 66,177; TRUE = 3,275		69,452	b
Cover_total	1-990	%	19,407	q
Cover_tree_layer	0.5-150	%	12,094	q
Cover_shrub_layer	0.5-170	%	16,804	q
Cover_herb_layer	0.2-199	%	29,668	q
Cover_moss_layer	1-100	%	9,681	q
Cover_lichen_layer	1-90	%	708	q
Cover_algae_layer	1-100	%	41	q
Cover_litter_layer	1-107	%	3,161	q
Cover_bare_rocks	1-100	%	2,747	q
Cover_cryptogams	1-90	%	772	q
Cover_bare_soil	0-99	%	2,746	q
Height_trees_highest	1-99	m	8,220	q
Height_trees_lowest	1-90	m	447	q
Height_shrubs_highest	0.1-9.9	m	3,389	q

(Continues)

TABLE 2 (Continued)

Variable	Range/levels	Unit of measurement	No. of plots with information	Type
Height_shrubs_lowest	0.1–9	m	263	q
Height_herbs_average	0.1–600	cm	5,901	q
Height_herbs_lowest	1–150	cm	490	q
Height_herbs_highest	1–600	cm	1,083	q
SoilClim_PC1	–6.233 – 8.172		95,104	q
SoilClim_PC2	–4.824 – 15.466		95,104	q
Resample_1	FALSE = 45,317; TRUE = 49,787		95,104	b
Resample_2	FALSE = 45,293; TRUE = 49,811		95,104	b
Resample_3	FALSE = 45,315; TRUE = 49,789		95,104	b
Resample_1_consensus	FALSE = 41,842; TRUE = 53,262		95,104	b

Note: dbh = diameter at breast height. Variable types can be n = nominal (i.e., qualitative variable); o = ordinal; q = quantitative; b = binary (i.e., Boolean); or d = date. Additional details on the variables are in Bruehlheide et al. (2019). Global Index of Vegetation Databases (GIVD) codes derive from Dengler et al. (2011). Biomes refer to Schultz (2005), modified to include also the world mountain regions (Körner et al., 2017). The column ESY refers to the European Nature Information System (EUNIS) Habitat Classification expert system (ESY, Chytrý et al., 2020).

2001). During this conversion, we checked that all datasets contained the required metadata information, and cross-checked that each plot was located within the geographical scope of its respective dataset. All individual TURBOVEG 2 datasets were then integrated into a TURBOVEG 3 database, and exported to comma-separated files. Finally, we harmonized all the taxonomic names from all datasets, based on sPlot's taxonomic backbone (Purschke, 2017). This backbone matched all the taxonomic names (without nomenclatural authors) from all datasets in sPlot v2.1 and TRY v3.0 (Kattge et al., 2020) to their resolved version based on the Taxonomic Name Resolution Service web application (TNRS version 4.0; Boyle et al., 2013). This allowed us to (a) harmonize all datasets to a common nomenclature and (b) link the sPlot database to the TRY database (Kattge et al., 2020). The final backbone only retained matched taxonomic names at the rank of species or higher. Additional detail on the taxonomic resolution is reported in Bruehlheide et al. (2019), while a description of the workflow, including R-code, is available in Purschke (2017).

6 | USAGE NOTES

The sPlotOpen database can be downloaded from <https://doi.org/10.25829/ivid.3474-40-3292>. A short vignette introducing the use of sPlotOpen in R can be found in Supporting Information Appendix S1. Users are urged to cite the original sources when using sPlotOpen in addition to the present paper (see Table 1). For two datasets (AF-00–009, AF-CD-001), the identification of taxa at species level is still in progress. Data on lichens and mosses, where available (e.g., dataset NA-GL-001), can be obtained on request from the respective dataset custodian or sPlot coordinator. As most of the constitutive datasets remain under continuous development, sPlotOpen users are encouraged to get in touch with the custodian(s) of the data they are planning to use (the updated list of custodian names is maintained on the sPlot website).

The use of sPlotOpen comes with a number of warnings. First, sPlotOpen was resampled in a way that maximizes the compositional variability of vegetation in different environmental conditions. As such, sPlotOpen should not be considered as representative of the spatial distribution of plant communities, especially when the focus has a local or regional spatial extent. Second, for most regions data were collected opportunistically, and without a randomized sampling design. This might lead to some vegetation types being oversampled in some regions, but undersampled in other regions, which might affect the output of species distribution models, especially at local or regional spatial extents. Third, not all plots were sampled using the same plot size, and some plots, mostly located in tropical regions, only contain data on woody species. This should be accounted for when exploring biodiversity patterns or comparing biodiversity indices (e.g., species richness, beta diversity) across plots or regions. Finally, a small fraction of plots are nested subsets of larger plots. Depending on the application, this might or might not represent a problem. Nested plots can be identified using the information in the ‘metadata’ matrix. The most appropriate way to deal with these issues depends on the problem being analysed. Users are, therefore, invited to carefully consider the limitations above when designing applications relying on sPlotOpen.

The data described here represent the subset of sPlot for which we were able to secure permission for making these data open. Additional data from sPlot are available under sPlot's Governance and Data Property Rules (<https://www.ivid.de/en/splot>). Using the full sPlot dataset is also recommended if a stratification is desired that is different from the environmental factors used here, for example by geographical region or plot size.

ACKNOWLEDGMENTS

The authors are grateful to the thousands of vegetation scientists who sampled vegetation plots in the field or digitized them into regional, national or international databases. The authors also

appreciate the support of the German Research Foundation for funding sPlot as one of the iDiv (DFG FZT 118, 202548816) research platforms, as well as for funding the position of Francesco Maria Sabatini and the organization of three workshops through the sDiv calls. The authors acknowledge this support with naming the database 'sPlot', where the 's' refers to the sDiv synthesis workshops. The authors are also grateful to Anahita Kazem and iDiv's Data & Code Unit for assistance with curation and archiving of the dataset.

The study has been supported by the TRY initiative on plant traits (<http://www.try-db.org>). The TRY initiative and database is hosted, developed and maintained by J. Kattge and G. Bönsch (Max Planck Institute for Biogeochemistry, Jena, Germany). TRY is currently supported by DIVERSITAS/Future Earth and iDiv Halle-Jena-Leipzig. Jens Kattge acknowledges support by the Max Planck Institute for Biogeochemistry (Jena, Germany), Future Earth, iDiv Halle-Jena-Leipzig and the EU H2020 project BACI, Grant No. 640176.

Isabelle Aubin was funded through the Natural Sciences and Engineering Research Council of Canada and Ontario Ministry of Natural Resources and Forestry. Yves Bergeron was funded through the Natural Sciences and Engineering Research Council of Canada. Idoia Biurrun was funded by the Basque Government (IT936-16). Anne Bjorkman thanks the Herschel Island-Qikiqtaruk Territorial Park management, Catherine Kennedy, Dorothy Cooley, Jill F. Johnstone, Cameron Eckert and Richard Gordon for establishing the ecological monitoring programme. Funding was provided by Herschel Island-Qikiqtaruk Territorial Park. Luis Cayuela was supported by project BIOCON08_044 funded by Fundación BBVA (Banco Bilbao Vizcaya Argentaria). Milan Chytrý, Flavia Landucci, Corrado Marcenò and Tomáš Peterka were supported by the Czech Science Foundation (project no. 19-28491X). Brian Enquist thanks the following individuals and institutions for contributing data to sPlot via the SALVIAS database: Mauricio Bonifacio, Saara DeWalt, Timothy Killeen, Susan Letcher, Nigel Pitman, Cam Webb, The Missouri Botanical Garden, RAINFOR and the Amazon Forest Inventory Network. Alvaro G. Gutiérrez was funded by Project FORECOFUN-SSA PIEF-GA-2010-274798 and FONDECYT 1200468. Mohamed Z. Hatim thanks Kamal Shaltout and Joop Schaminée for MSc thesis supervision, and Joop Schaminée for support and funding from the Prince Bernard Culture Fund Prize for Nature Conservation. Jürgen Homeier received funding from BMBF (Federal Ministry of Education and Science of Germany) and the German Research Foundation (DFG Ho3296-2, DFG Ho3296-4). Borja Jiménez-Alfaro was funded by the Spanish Research Agency through grant AEI/10.13039/501100011033. Dirk N. Karger received funding from: the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL) internal grant exCHELSA and ClimEx, the Joint Biodiversa COFUND project 'FeedBacks' and 'Futureweb', the Swiss Data Science Projects: SPEEDMIND, and COMECO, and the Swiss National Science Foundation (20BD21_184131). Hjalmar Kühl gratefully acknowledges the Pan African team and funding by the Max Planck Society and Krekeler Foundation. Attila

Lengyel was supported by the National Research, Development and Innovation Office, Hungary (PD-123997). Tatiana Lysenko was funded by the Russian Foundation for Basic Research (Grant No. 16-04-00747a). Alireza Naqinezhad is supported by a master grant from the University of Mazandaran. Jérôme Munzinger was supported by the French National Research Agency (ANR) with grants INC (ANR-07-BDIV-0008), BIONEOCAL (ANR-07-BDIV-0006) & ULTRABIO (ANR-07-BDIV-0010), by the National Geographic Society (Grant 7579-04), and with funding and authorizations of North and South Provinces of New Caledonia. Arkadiusz Nowak received support from the National Science Centre, Poland, grant no. 2017/25/B/NZ8/00572. Gerhard E. Overbeck acknowledges support from Brazil's National Council of Scientific and Technological Development (CNPq, grant 310022/2015-0). Meelis Pärtel was supported by the Estonian Research Council (PRG609) and European Regional Development Fund (Centre of Excellence EcolChange). Robert Peet acknowledges the support from the National Center for Ecological Analysis and Synthesis, the North Carolina Ecosystem Enhancement Program, the U.S. Forest Service, and the U.S. National Science Foundation (DBI-9905838, DBI-0213794). Josep Peñuelas acknowledges the financial support from the European Research Council Synergy grant ERC-SyG-2013-610028 IMBALANCE-P. Petr Petřík and Jiri Dolezal acknowledge the support of the long-term research development project No. RVO 67985939 of the Czech Academy of Sciences. Oliver Phillips was funded by an ERC Advanced Grant (291585, 'T-FORCES') and a Royal Society-Wolfson Research Merit Award. Valério D. Pillar was supported by the Brazil's National Council of Scientific and Technological Development (CNPq, grant 307689/2014-0). Solvita Rūsiņa was supported by the University of Latvia grant AAP2016/B041/Zd2016/AZ03 within the 'Climate change and sustainable use of natural resources' framework. Franziska Schrodtt was supported by the University of Minnesota Institute on the Environment Discovery Grant, the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig grant (50170649_#7) and the University of Nottingham Anne McLaren Fellowship. Jozef Šibík was funded by The Slovak Research and Development Agency grant no. APVV16-0431. Jens-Christian Svenning considers this work a contribution to his VILLUM Investigator project 'Biodiversity Dynamics in a Changing World' funded by VILLUM FONDEN (grant 16549). Kim André Vanselow would like to thank W. Bernhard Dickoré for the help in the identification of plant species and acknowledges the financial support from the Volkswagen Foundation (AZ I/81 976) and the German Research Foundation (DFG VA 749/1-1, DFG VA 749/4-1). Evan Weiher was funded by NSF DEB-0415383, UWEC-ORSP, and UWEC-BCDT. Work by Karsten Wesche was supported by the German Research Foundation (DFG WE 2601/3-1,3-2, 4-1,4-2) and by the German Ministry for Science and Education (BMBF, CAME 03G0808A). Susan Wiser was funded by the New Zealand (NZ) Ministry for Business, Innovation and Employment's Strategic Science Investment Fund.

This paper is dedicated to the memory of Dr. Ching-Feng (Woody) Li.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

FMS wrote the first draft of the manuscript, with considerable input from JL and HB. JL and TH wrote the resampling algorithm. FMS set up the GitHub projects, curated the database, and produced the graphs. He also coordinated the sPlot consortium. SMH wrote the TURBOVEG software, which holds the sPlot database. JKa provided the trait data from TRY and FSc performed the trait data gap filling. HB secured the funding for sPlot as a strategic project of iDiv. All other authors contributed data and/or helped set up the database and/or helped develop the resampling algorithm. All authors contributed to revising and approved the manuscript.

DATA AVAILABILITY STATEMENT

The R code used to produce sPlotOpen from the sPlot v2.1 database is contained in the *sPlotOpen_code* GitHub repository: https://github.com/fmsabatini/sPlotOpen_Code. This manuscript was produced using the Manubot workflow (Himmelstein et al., 2019). The code for reproducing this manuscript is stored in the *sPlotOpen_manuscript* GitHub repository: https://github.com/fmsabatini/sPlotOpen_Manuscript.

ORCID

Francesco Maria Sabatini  <https://orcid.org/0000-0002-7202-7697>
Jonathan Lenoir  <https://orcid.org/0000-0003-0638-9582>
Tarek Hattab  <https://orcid.org/0000-0002-1420-5758>
Elise Aimee Arnst  <https://orcid.org/0000-0003-2388-7428>
Milan Chytrý  <https://orcid.org/0000-0002-8122-3075>
Jürgen Dengler  <https://orcid.org/0000-0003-3221-660X>
Stephan M. Hennekens  <https://orcid.org/0000-0003-1221-0323>
Ute Jandt  <https://orcid.org/0000-0002-3177-3669>
Florian Jansen  <https://orcid.org/0000-0002-0331-5185>
Borja Jiménez-Alfaro  <https://orcid.org/0000-0001-6601-9597>
Jens Kattge  <https://orcid.org/0000-0002-1022-8469>
Valério D. Pillar  <https://orcid.org/0000-0001-6408-2891>
Oliver Purschke  <https://orcid.org/0000-0003-0444-0882>
Brody Sandel  <https://orcid.org/0000-0003-2162-6902>
Tsipe Aavik  <https://orcid.org/0000-0001-5232-3950>
Svetlana Ačić  <https://orcid.org/0000-0001-6553-3797>
Alicia T. R. Acosta  <https://orcid.org/0000-0001-6572-3187>
Emiliano Agrillo  <https://orcid.org/0000-0003-2346-8346>
Miguel Alvarez  <https://orcid.org/0000-0003-1500-1834>
Mohammed A. S. Arfin Khan  <https://orcid.org/0000-0001-6275-7023>
Fabio Attorre  <https://orcid.org/0000-0002-7744-2195>
Isabelle Aubin  <https://orcid.org/0000-0002-5953-1012>
Marijn Bauters  <https://orcid.org/0000-0003-0978-6639>
Yves Bergeron  <https://orcid.org/0000-0003-3707-3687>
Erwin Bergmeier  <https://orcid.org/0000-0002-6118-4611>
Idoia Biurrun  <https://orcid.org/0000-0002-1454-0433>
Anne D. Bjorkman  <https://orcid.org/0000-0003-2174-7800>
Gianmaria Bonari <https://orcid.org/0000-0002-5574-6067>

Viktoria Bondareva  <https://orcid.org/0000-0002-6676-5722>
Jörg Brunet  <https://orcid.org/0000-0003-2667-4575>
Andraž Čarni  <https://orcid.org/0000-0002-8909-4298>
Laura Casella  <https://orcid.org/0000-0003-2550-3010>
Luis Cayuela  <https://orcid.org/0000-0003-3562-2662>
Victor Chepinoga  <https://orcid.org/0000-0003-3809-7453>
János Csiky  <https://orcid.org/0000-0002-7920-5070>
Els De Bie  <https://orcid.org/0000-0001-7679-743X>
André Luis de Gasper  <https://orcid.org/0000-0002-1940-9581>
Michele De Sanctis  <https://orcid.org/0000-0002-7280-6199>
Jiri Dolezal  <https://orcid.org/0000-0002-5829-4051>
Mohamed Abd El-Rouf Mousa El-Sheikh  <https://orcid.org/0000-0002-0720-7448>
Brian Enquist  <https://orcid.org/0000-0002-6124-7096>
Jörg Ewald  <https://orcid.org/0000-0002-2758-9324>
Richard Field  <https://orcid.org/0000-0003-2613-2688>
Manfred Finckh  <https://orcid.org/0000-0003-2186-0854>
Sophie Gachet  <https://orcid.org/0000-0002-3599-5189>
Antonio Galán-de-Mera  <https://orcid.org/0000-0002-1652-5931>
Hamid Gholizadeh  <https://orcid.org/0000-0002-3694-368X>
Melisa Giorgis  <https://orcid.org/0000-0001-6126-6660>
Valentin Golub  <https://orcid.org/0000-0003-3973-6608>
Inger Greve Alsos  <https://orcid.org/0000-0002-8610-1085>
Gregory Richard Guerin  <https://orcid.org/0000-0002-2104-6695>
Alvaro G. Gutiérrez  <https://orcid.org/0000-0001-8928-3198>
Sylvia Haider  <https://orcid.org/0000-0002-2966-0534>
Mohamed Z. Hatim  <https://orcid.org/0000-0002-0872-5108>
Bruno Hérault  <https://orcid.org/0000-0002-6950-7286>
Norbert Hölzel  <https://orcid.org/0000-0002-6367-3400>
Jürgen Homeier  <https://orcid.org/0000-0001-5676-3267>
Anke Jentsch  <https://orcid.org/0000-0002-2345-8300>
Norbert Jürgens  <https://orcid.org/0000-0003-3211-0549>
Dirk Nikolaus Karger  <https://orcid.org/0000-0001-7770-6229>
Ali Kavğacı  <https://orcid.org/0000-0002-4549-3668>
Elizabeth Kearsley  <https://orcid.org/0000-0003-0046-3606>
Michael Kessler  <https://orcid.org/0000-0003-4612-9937>
Larisa Khanina  <https://orcid.org/0000-0002-8937-5938>
Holger Kreft  <https://orcid.org/0000-0003-4471-8236>
Hjalmar S. Kühl  <https://orcid.org/0000-0002-4440-9161>
Anna Kuzemko  <https://orcid.org/0000-0002-9425-2756>
Flavia Landucci  <https://orcid.org/0000-0002-6848-0384>
Attila Lengyel  <https://orcid.org/0000-0002-1712-6748>
Frederic Lens  <https://orcid.org/0000-0002-5001-0149>
Débora Vanessa Lingner  <https://orcid.org/0000-0002-6391-9343>
Hongyan Liu  <https://orcid.org/0000-0002-6721-4439>
Tatiana Lysenko  <https://orcid.org/0000-0001-6688-1590>
Miguel D. Mahecha  <https://orcid.org/0000-0003-3031-613X>
Corrado Marcenò  <https://orcid.org/0000-0003-4361-5200>
Vasiliy Martynenko  <https://orcid.org/0000-0002-9071-3789>
Jesper Erenskjold Moeslund  <https://orcid.org/0000-0001-8591-7149>
Abel Monteagudo Mendoza  <https://orcid.org/0000-0002-1047-845X>
Ladislav Mucina <https://orcid.org/0000-0003-0317-8886>
Jonas V. Müller <https://orcid.org/0000-0001-7049-3048>

Jérôme Munzinger  <https://orcid.org/0000-0001-5300-2702>
 Jalil Noroozi  <https://orcid.org/0000-0003-4124-2359>
 Arkadiusz Nowak  <https://orcid.org/0000-0001-8638-0208>
 Gerhard E. Overbeck  <https://orcid.org/0000-0002-8716-5136>
 Meelis Pärtel  <https://orcid.org/0000-0002-5874-0138>
 Aníbal Pauchard  <https://orcid.org/0000-0003-1284-3163>
 Robert K. Peet  <https://orcid.org/0000-0003-2823-6587>
 Josep Peñuelas  <https://orcid.org/0000-0002-7215-0150>
 Aaron Pérez-Haase  <https://orcid.org/0000-0002-5974-7374>
 Petr Petřík  <https://orcid.org/0000-0001-8518-6737>
 Gwendolyn Peyre  <https://orcid.org/0000-0002-1977-7181>
 Oliver L. Phillips  <https://orcid.org/0000-0002-8993-6168>
 Valerijus Rašomavičius  <https://orcid.org/0000-0003-1314-4356>
 Rasmus Revermann  <https://orcid.org/0000-0002-7044-768X>
 Gonzalo Rivas-Torres  <https://orcid.org/0000-0002-2704-8288>
 Solvita Rūsiņa  <https://orcid.org/0000-0002-9580-4110>
 Marco Schmidt  <https://orcid.org/0000-0001-6087-6117>
 Franziska Schrodtt  <https://orcid.org/0000-0001-9053-8872>
 Pavel Shirokikh  <https://orcid.org/0000-0003-1864-4878>
 Jozef Šibík  <https://orcid.org/0000-0002-5949-862X>
 Urban Šilc  <https://orcid.org/0000-0002-3052-699X>
 Željko Škvorc  <https://orcid.org/0000-0002-2848-1454>
 Marta Gaia Sperandii  <https://orcid.org/0000-0002-2507-5928>
 Jens-Christian Svenning  <https://orcid.org/0000-0002-3415-0862>
 Kim André Vanselow  <https://orcid.org/0000-0003-3299-6220>
 Eduardo Vélez-Martin  <https://orcid.org/0000-0001-8028-8953>
 Roberto Venanzoni  <https://orcid.org/0000-0002-7768-0468>
 Cyrille Violle  <https://orcid.org/0000-0002-2471-9226>
 Risto Virtanen  <https://orcid.org/0000-0002-8295-8217>
 Viktoria Wagner  <https://orcid.org/0000-0002-2665-9888>
 Donald A. Walker  <https://orcid.org/0000-0002-2665-9888>
 Donald M. Waller  <https://orcid.org/0000-0001-5377-3929>
 Timothy J. S. Whitfeld  <https://orcid.org/0000-0003-1850-6432>
 Wolfgang Willner  <https://orcid.org/0000-0003-1591-8386>
 Susan K. Wiser  <https://orcid.org/0000-0002-8938-8181>
 Thomas Wohlgemuth  <https://orcid.org/0000-0002-4623-0894>
 Martin Zobel  <https://orcid.org/0000-0001-7957-6704>
 Helge Bruelheide  <https://orcid.org/0000-0003-3135-0356>

REFERENCES

- Ačić, S., Petrović, M., Dajić Stevanović, Z., & Šilc, U. (2012). Vegetation Database Grassland Vegetation in Serbia. *Biodiversity & Ecology*, 4, 418.
- Agrillo, E., Alessi, N., Massimi, M., Spada, F., De Sanctis, M., Francesconi, F., & Attorre, F. (2017). Nationwide vegetation plot database—Sapienza University of Rome: State of the art, basic figures and future perspectives. *Phytocoenologia*, 47, 221–229.
- Alvarez, M., Curran, M., & Malombe, I. (2021). SWEA-Dataveg: A vegetation database for sub-Saharan Africa. *Vegetation Classification and Survey*, 2, 59–63. <https://doi.org/10.3897/VCS/2021/64911>
- Apostolova, I., Sopotlieva, D., Pedashenko, H., Velev, N., & Vasilev, K. (2012). Bulgarian Vegetation Database: Historic background, current status and future prospects. *Biodiversity & Ecology*, 4, 141–148. <https://doi.org/10.7809/b-e.00069>
- Aubin, I., Cardou, F., Boisvert-Marsh, L., Garnier, E., Strukelj, M., & Munson, A. D. (2020). Managing data locally to answer questions globally: The role of collaborative science in ecology. *Journal of Vegetation Science*, 31, 509–517. <https://doi.org/10.1111/jvs.12864>
- Aubin, I., Gachet, S., Messier, C., & Bouchard, A. (2007). How resilient are northern hardwood forests to human disturbance? An evaluation using a plant functional group approach. *Ecoscience*, 14, 259–271.
- Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences USA*, 115, 6506–6511. <https://doi.org/10.1073/pnas.1711842115>
- Beck, J. J., Larget, B., & Waller, D. M. (2018). Phantom species: Adjusting estimates of colonization and extinction for pseudo-turnover. *Oikos*, 127, 1605–1618. <https://doi.org/10.1111/oik.05114>
- Biurrun, I., García-Mijangos, I., Campos, J. A., Herrera, M., & Loidi, J. (2012). Vegetation-plot database of the University of the Basque Country (BIOVEG). *Biodiversity & Ecology*, 4, 328. <https://doi.org/10.7809/b-e.00121>
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, 8, e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Bonebrake, T. C., Brown, C. J., Bell, J. D., Blanchard, J. L., Chauvenet, A., Champion, C., Chen, I.-C., Clark, T. B., Colwell, R. K., Danielsen, F., Dell, A. I., Donelson, J. M., Evengård, B., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Jarzyna, M. A., ... Pecl, G. T. (2018). Managing consequences of climate-driven species redistribution requires integration of ecology, conservation and social science. *Biological Reviews*, 93, 284–305. <https://doi.org/10.1111/brv.12344>
- Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J. A., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., McKay, S. J., Lowry, S., Freeland, C., Peet, R. K., & Enquist, B. J. (2013). The taxonomic name resolution service: An online tool for automated standardization of plant names. *BMC Bioinformatics*, 14, 1–16. <https://doi.org/10.1186/1471-2105-14-16>
- Bruelheide, H., Böhne, M., Both, S., Fang, T., Assmann, T., Baruffol, M., Bauhus, J., Buscot, F., Chen, X.-Y., Ding, B.-Y., Durka, W., Erfmeier, A., Fischer, M., Geißler, C., Guo, D., Guo, L.-D., Härdtle, W., He, J.-S., Hector, A., ... Schmid, B. (2011). Community assembly during secondary forest succession in a Chinese subtropical forest. *Ecological Monographs*, 81, 25–41. <https://doi.org/10.1890/09-2172.1>
- Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., Pillar, V. D., Jansen, F., Kattge, J., Sandel, B., Aubin, I., Biurrun, I., Field, R., Haider, S., Jandt, U., Lenoir, J., Peet, R. K., Peyre, G., Sabatini, F. M., ... Zverev, A. (2019). sPlot—A new tool for global vegetation analyses. *Journal of Vegetation Science*, 30, 161–186. <https://doi.org/10.1111/jvs.12710>
- Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S. M., Botta-Dukát, Z., Chytrý, M., Field, R., Jansen, F., Kattge, J., Pillar, V. D., Schrodtt, F., Mahecha, M. D., Peet, R. K., Sandel, B., van Bodegom, P., Altman, J., Alvarez-Dávila, E., ... Jandt, U. (2018). Global trait–environment relationships of plant communities. *Nature Ecology & Evolution*, 2, 1906–1917. <https://doi.org/10.1038/s41559-018-0699-8>
- Cai, Q., Welk, E., Ji, C., Fang, W., Sabatini, F. M., Zhu, J., Zhu, J., Tang, Z., Attorre, F., Campos, J. A., Čarni, A., Chytrý, M., Çoban, S., Dengler, J., Dolezal, J., Field, R., Frink, J. P., Gholizadeh, H., Indreica, A., ... Bruelheide, H. (2021). The relationship between niche breadth and range size of beech (*Fagus*) species worldwide. *Journal of Biogeography*, 48, 1240–1253. <https://doi.org/10.1111/jbi.14074>
- Cardinale, B. J., Matulich, K. L., Hooper, D. U., Byrnes, J. E., Duffy, E., Gamfeldt, L., Balvanera, P., O'Connor, M. I., & Gonzalez, A. (2011). The functional role of producer diversity in ecosystems. *American Journal of Botany*, 98, 572–592. <https://doi.org/10.3732/ajb.1000364>
- Casella, L., Bianco, P. M., Angelini, P., & Morroni, E. (2012). Italian national vegetation database (BVN/ISPRA). *Biodiversity & Ecology*, 4, 404. <https://doi.org/10.7809/b-e.00192>

- Cayuela, L., Gálvez-Bravo, L., Pérez Pérez, R., de Albuquerque, F., Golicher, D., Zahawi, R., Ramírez-Marcial, N., Garibaldi, C., Field, R., Rey Benayas, J., González-Espinosa, M., Balvanera, P., Ángel Castillo, M., Figueroa-Rangel, B., Griffith, D., Islebe, G., Kelly, D., Olvera-Vargas, M., Schnitzer, S., ... Zamora, R. (2012). The tree biodiversity network (BIOTREE-NET): Prospects for biodiversity research and conservation in the Neotropics. *Biodiversity & Ecology*, 4, 211–224. <https://doi.org/10.7809/b-e.00078>
- Černý, T., Kopecký, M., Petřík, P., Song, J.-S., Šrůtek, M., Valachovič, M., Altman, J., & Doležal, J. (2015). Classification of Korean forests: Patterns along geographic and environmental gradients. *Applied Vegetation Science*, 18, 5–22. <https://doi.org/10.1111/avsc.12124>
- Chabbi, A., & Loescher, H. W. (2017). *Terrestrial ecosystem research infrastructures: Challenges and opportunities*. CRC Press.
- Chepinoga, V. V. (2012). Wetland Vegetation Database of Baikal Siberia (WETBS). *Biodiversity & Ecology*, 4, 311. <https://doi.org/10.7809/b-e.00107>
- Chytrý, M. (2012). Database of Masaryk University vegetation research in Siberia. *Biodiversity & Ecology*, 4, 290.
- Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J. H. J., Ačić, S., Agrillo, E., Ambarli, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Brisse, H., ... Yamalov, S. (2016). European Vegetation Archive (EVA): An integrated database of European vegetation plots. *Applied Vegetation Science*, 19, 173–180. <https://doi.org/10.1111/avsc.12191>
- Chytrý, M., & Rafajová, M. (2003). Czech National Phytosociological Database: Basic statistics of the available vegetation-plot data. *Preslia*, 75, 1–15.
- Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J. S., Peterka, T., Marcenò, C., Landucci, F., Danihelka, J., Hájek, M., Dengler, J., Novák, P., Zúkal, D., Jiménez-Alfaro, B., Mucina, L., Abdulhak, S., Ačić, S., Agrillo, E., ... Schaminée, J. H. J. (2020). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, 23, 648–675. <https://doi.org/10.1111/avsc.12519>
- De Sanctis, M., & Attorre, F. (2012). Socotra Vegetation Database. *Biodiversity & Ecology*, 4, 315. <https://doi.org/10.7809/b-e.00111>
- De Sanctis, M., Fanelli, G., Mullaj, A., & Attorre, F. (2017). Vegetation database of Albania. *Phytocoenologia*, 47, 107–108. <https://doi.org/10.1127/phyto/2017/0178>
- Dengler, J., Jansen, F., Glöckler, F., Peet, R. K., De Cáceres, M., Chytrý, M., Ewald, J., Oldeland, J., Lopez-Gonzalez, G., Finckh, M., Mucina, L., Rodwell, J. S., Schaminée, J. H. J., & Spencer, N. (2011). The Global Index of Vegetation-Plot Databases (GIVD): A new resource for vegetation science. *Journal of Vegetation Science*, 22, 582–597. <https://doi.org/10.1111/j.1654-1103.2011.01265.x>
- Dengler, J., & Rüşina, S. (2012). Database dry grasslands in the Nordic and Baltic region. *Biodiversity & Ecology*, 4, 319–320. <https://doi.org/10.7809/b-e.00114>
- Dimopoulos, P., & Tsiripidis, I. (2012). Hellenic Natura 2000 Vegetation Database (HelNatVeg). *Biodiversity & Ecology*, 4, 388. <https://doi.org/10.7809/b-e.00177>
- Elmendorf, S. C., Henry, G. H. R., Hollister, R. D., Björk, R. G., Boulanger-Lapointe, N., Cooper, E. J., & Wipf, S. (2012). Plot-scale evidence of tundra vegetation change and links to summer warming. *Nature Climate Change*, 2, 453–457.
- El-Sheikh, M. A., Thomas, J., Alfarhan, A. H., Alatar, A. A., Mayandy, S., Hennekens, S. M., Schaminée, J. H. J., Mucina, L., & Alansari, A. M. (2017). SaudiVeg ecoinformatics: Aims, current status and perspectives. *Saudi Journal of Biological Sciences*, 24, 389–398. <https://doi.org/10.1016/j.sjbs.2016.02.012>
- Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M., & Thiers, B. M. (2016). Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ*, 4, e2615v2.
- Enquist, B. J., Norberg, J., Bonser, S. P., Violle, C., Webb, C. T., Henderson, A., Sloat, L. L., & Savage, V. M. (2015). Scaling from traits to ecosystems: Developing a general trait driver theory via integrating trait-based and metabolic scaling theories. In S. Pawar, G. Woodward, & A. I. Dell (Eds.), *Advances in ecological research* (pp. 249–318). Academic Press.
- Ewald, J., May, R., & Kleikamp, M. (2012). VegetWeb—the national online-repository of vegetation plots from Germany. *Biodiversity & Ecology*, 4, 173–175. <https://doi.org/10.7809/b-e.00073>
- Fazayeli, F., Banerjee, A., Kattge, J., Schrödt, F., & Reich, P. B. (2014). Uncertainty quantified matrix completion using Bayesian hierarchical matrix factorization. In 2014 13th international conference on machine learning and applications (pp. 312–317). <https://doi.org/10.1109/ICMLA.2014.56>
- Finckh, M. (2012). Vegetation database of southern Morocco. *Biodiversity & Ecology*, 4, 297. <https://doi.org/10.7809/b-e.00094>
- Fotiadis, G., Tsiripidis, I., Bergmeier, E., & Dimopoulos, P. (2012). Hellenic woodland database. *Biodiversity & Ecology*, 4, 389. <https://doi.org/10.7809/b-e.00178>
- Franklin, J., Serra-Díaz, J. M., Syphard, A. D., & Regan, H. M. (2017). Big data for forecasting the impacts of global change on plant communities. *Global Ecology and Biogeography*, 26, 6–17. <https://doi.org/10.1111/geb.12501>
- Fricke, E. C., & Svenning, J.-C. (2020). Accelerating homogenization of the global plant–frugivore meta-network. *Nature*, 585, 74–78. <https://doi.org/10.1038/s41586-020-2640-y>
- Garbolino, E., De Ruffray, P., Brisse, H., & Grandjouan, G. (2012). The phytosociological database SOPHY as the basis of plant socio-ecology and phytoclimatology in France. *Biodiversity & Ecology*, 4, 177–184. <https://doi.org/10.7809/b-e.00074>
- Golub, V., Sorokin, A., Starichkova, K., Nikolaychuk, L., Bondareva, V., & Ivakhnova, T. (2012). Lower Volga valley phytosociological database. *Biodiversity & Ecology*, 4, 589.
- Harper, K., Boudreault, C., DeGrandpré, L., Drapeau, P., Gauthier, S., & Bergeron, Y. (2003). Structure, composition, and diversity of old-growth black spruce boreal forest of the Clay Belt region in Quebec and Ontario. *Environmental Reviews*, 11, S79–S98. <https://doi.org/10.1139/a03-013>
- Hatim, M. (2012). Vegetation database of Sinai in Egypt. *Biodiversity & Ecology*, 4, 303. <https://doi.org/10.7809/b-e.00099>
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12, e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Hennekens, S. M., & Schaminée, J. H. J. (2001). TURBOVEG, a comprehensive data base management system for vegetation data. *Journal of Vegetation Science*, 12, 589–591. <https://doi.org/10.2307/3237010>
- Himmelstein, D. S., Rubinetti, V., Slochower, D. R., Hu, D., Malladi, V. S., Greene, C. S., & Gitter, A. (2019). Open collaborative writing with Manubot. *PLoS Computational Biology*, 15, e1007128. <https://doi.org/10.1371/journal.pcbi.1007128>
- Ibanez, T., Munzinger, J., Dagostini, G., Hequet, V., Rigault, F., Jaffré, T., & Birnbaum, P. (2014). Structural and floristic diversity of mixed rainforest in New Caledonia: New data from the New Caledonian Plant Inventory and Permanent Plot Network (NC-PIPPN). *Applied Vegetation Science*, 17, 386–397.
- Indreica, A., Turtureanu, P. D., Szabó, A., & Irimia, I. (2017). Romanian forest database: A phytosociological archive of woody vegetation. *Phytocoenologia*, 47, 389–393. <https://doi.org/10.1127/phyto/2017/0201>

- IPBES (Ed.) (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES Secretariat.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. 1. *New Phytologist*, 11, 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Jandt, U., & Bruehlheide, H. (2012). German Vegetation Reference Database (GVRD). *Biodiversity & Ecology*, 4, 355. <https://doi.org/10.7809/b-e.00146>
- Jandt, U., von Wehrden, H., & Bruehlheide, H. (2011). Exploring large vegetation databases to detect temporal trends in species occurrences. *Journal of Vegetation Science*, 22, 957–972. <https://doi.org/10.1111/j.1654-1103.2011.01318.x>
- Jansen, F., Dengler, J., & Berg, C. (2012). VegMV—the vegetation database of Mecklenburg-Vorpommern. *Biodiversity & Ecology*, 4, 149–160. <https://doi.org/10.7809/b-e.00070>
- Kącki, Z., & Śliwiński, M. (2012). The polish vegetation database: Structure, resources and development. *Acta Societatis Botanicorum Poloniae*, 81, 75–79. <https://doi.org/10.5586/asbp.2012.014>
- Karger, D. N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122. <https://doi.org/10.1038/sdata.2017.122>
- Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar, C. C., Aleixo, I., Ali, H., ... Wirth, C. (2020). TRY plant trait database—enhanced coverage and open access. *Global Change Biology*, 26, 119–188. <https://doi.org/10.1111/gcb.14904>
- Kearsley, E., de Haulleville, T., Hufkens, K., Kidimbu, A., Toirambe, B., Baert, G., Huygens, D., Kebede, Y., Defourny, P., Bogaert, J., Beeckman, H., Steppe, K., Boeckx, P., & Verbeeck, H. (2013). Conventional tree height-diameter relationships significantly overestimate aboveground carbon stocks in the Central Congo Basin. *Nature Communications*, 4, 2269. <https://doi.org/10.1038/ncomm3269>
- Körner, C., Jetz, W., Paulsen, J., Payne, D., Rudmann-Maurer, K., & Spehn, M. E. (2017). A global inventory of mountains for bio-geographical applications. *Alpine Botany*, 127, 1–15. <https://doi.org/10.1007/s00035-016-0182-6>
- Korolyuk, A. Y., & Zverev, A. (2012). Database of Siberian Vegetation (DSV). *Biodiversity & Ecology*, 4, 312. <https://doi.org/10.7809/b-e.00108>
- Kühl, H. S., Bowler, D. E., Bösch, L., Bruehlheide, H., Dauber, J., Eichenberg, D., Eisenhauer, N., Fernández, N., Guerra, C. A., Henle, K., Herbing, I., Isaac, N. J. B., Jansen, F., König-Ries, B., Kühn, I., Nilsen, E. B., Pe'er, G., Richter, A., Schulte, R., ... Bonn, A. (2020). Effective biodiversity monitoring needs a culture of integration. *One Earth*, 3, 462–474. <https://doi.org/10.1016/j.oneear.2020.09.010>
- Kuzemko, A. (2012). Ukrainian grasslands database. *Biodiversity & Ecology*, 4, 430. <https://doi.org/10.7809/b-e.00217>
- Lájer, K., Botta-Dukát, Z., Csiky, J., Horváth, F., Szmorad, F., Bagi, I., Dobolyi, Z. K., Hahn, I., Kovács, J. A., & Rédei, T. (2008). Hungarian phytosociological database (COENODATREF): Sampling methodology, nomenclature and its actual stage. *Annali Di Botanica, Nuova Serie*, 7, 197–201.
- Landucci, F., Acosta, A. T. R., Agrillo, E., Attorre, F., Biondi, E., Cambria, V. E., Chiarucci, A., Del Vico, E., De Sanctis, M., Facioni, L., Geri, F., Gigante, D., Guarino, R., Landi, S., Lucarini, D., Panfili, E., Pesaresi, S., Prisco, I., Rosati, L., ... Venzonzi, R. (2012). VegItaly: The Italian collaborative project for a national vegetation database. *Plant Biosystems*, 146, 756–763. <https://doi.org/10.1080/11263504.2012.740093>
- Landucci, F., Řezníčková, M., Šumberová, K., Chytrý, M., Aunina, L., Biřá-Nicolae, C., Bobrov, A., Borsukevych, L., Brisse, H., Čarni, A., Csiky, J., Cvijanović, D., De Bie, E., De Ruffray, P., Dubyna, D., Dimopoulos, P., Dziuba, T., FitzPatrick, Ú., Font, X., ... Willner, W. (2015). WetVegEurope: A database of aquatic and wetland vegetation of Europe. *Phytocoenologia*, 45, 187–194. <https://doi.org/10.1127/phyto/2015/0050>
- Lengyel, A., Chytrý, M., & Tichý, L. (2011). Heterogeneity-constrained random resampling of phytosociological databases. *Journal of Vegetation Science*, 22, 175–183. <https://doi.org/10.1111/j.1654-1103.2010.01225.x>
- Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Muriene, J., & Grenouillet, G. (2020). Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*, 4, 1044–1059. <https://doi.org/10.1038/s41559-020-1198-2>
- Lenoir, J., Gégout, J.-C., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N. E., Dullinger, S., Pauli, H., Willner, W., Grytnes, J.-A., Virtanen, R., & Svenning, J.-C. (2010). Cross-scale analysis of the region effect on vascular plant species diversity in southern and northern European mountain ranges. *PLoS ONE*, 5, e15734. <https://doi.org/10.1371/journal.pone.0015734>
- Lenoir, J., Gégout, J. C., Marquet, P. A., de Ruffray, P., & Brisse, H. (2008). A significant upward shift in plant species optimum elevation during the 20th century. *Science*, 320, 1768–1771. <https://doi.org/10.1126/science.1156831>
- Lenoir, J., Graae, B. J., Aarrestad, P. A., Alsos, I. G., Armbruster, W. S., Austrheim, G., Bergendorff, C., Birks, H. J. B., Bråthen, K. A., Brunet, J., Bruun, H. H., Dahlberg, C. J., Decocq, G., Diekmann, M., Dynesius, M., Ejrnaes, R., Grytnes, J.-A., Hylander, K., Klanderud, K., ... Svenning, J.-C. (2013). Local temperatures inferred from plant communities suggest strong spatial buffering of climate warming across Northern Europe. *Global Change Biology*, 19, 1470–1481. <https://doi.org/10.1111/gcb.12129>
- Lenoir, J., & Svenning, J.-C. (2015). Climate-related range shifts—a global multidimensional synthesis and new research directions. *Ecography*, 38, 15–28. <https://doi.org/10.1111/ecog.00967>
- Liu, H., Cui, H., Pott, R., & Speier, M. (2000). Vegetation of the woodland-steppe ecotone in southeastern Inner Mongolia, China. *Journal of Vegetation Science*, 11, 525–532.
- Lopez-Gonzalez, G., Lewis, S. L., Burkitt, M., & Phillips, O. L. (2011). ForestPlots.net: A web application and research tool to manage and analyse tropical forest plot data. *Journal of Vegetation Science*, 22, 610–613. <https://doi.org/10.1111/j.1654-1103.2011.01312.x>
- Lysenko, T., Kalmykova, O., & Mitroshenkova, A. (2012). Vegetation database of the Volga and the Ural rivers basins. *Biodiversity & Ecology*, 4, 420–421. <https://doi.org/10.7809/b-e.00208>
- Marcenò, C., & Jiménez-Alfaro, B. (2017). The Mediterranean Ammophilete Database: A comprehensive dataset of coastal dune vegetation. *Phytocoenologia*, 47, 95–105.
- Muche, G., Schmiedel, U., & Jürgens, N. (2012). BIOTA Southern Africa Biodiversity Observatories Vegetation Database. *Biodiversity & Ecology*, 4, 111–123. <https://doi.org/10.7809/b-e.00066>
- Müller, J. (2003). *Zur Vegetationsökologie der Savannenlandschaften im Sahel Burkina Faso*. Frankfurt am Main: Goethe Universität Frankfurt am Main.
- Nowak, A., Nobis, M., Nowak, S., Nobis, A., Swacha, G., & Kącki, Z. (2017). Vegetation of Middle Asia: The project state of the art after ten years of survey and future perspectives. *Phytocoenologia*, 47, 395–400.
- Pauchard, A., Fuentes, N., Jiménez, A., Bustamante, R., & Marticorena, A. (2013). Alien plants homogenise protected areas: Evidence from the landscape and regional scales in south central Chile. In L. C. Foxcroft, P. Pyšek, D. M. Richardson, & G. Piero (Eds.), *Plant invasions in protected areas* (pp. 191–208). Springer. <https://doi.org/10.1007/978-94-007-7750-7>
- Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffiths, R. B., Hobday, A. J.,

- Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., ... Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355, eaai9214. <https://doi.org/10.1126/science.aai9214>
- Peet, R. K., Lee, M. T., Boyle, M. F., Wentworth, T. R., Schafale, M. P., & Weakley, A. S. (2012). Vegetation-plot database of the Carolina Vegetation Survey. *Biodiversity & Ecology*, 4, 243–253. <https://doi.org/10.7809/b-e.00081>
- Peet, R. K., Lee, M. T., Jennings, M. D., & Faber-Langendoen, D. (2012). VegBank—a permanent, open-access archive for vegetation-plot data. *Biodiversity and Ecology*, 4, 233–241. <https://doi.org/10.7809/b-e.00080>
- Perring, M. P., Bernhardt-Römermann, M., Baeten, L., Midolo, G., Blondeel, H., Depauw, L., Landuyt, D., Maes, S. L., De Lombaerde, E., Carón, M. M., Vellend, M., Brunet, J., Chudomelová, M., Decocq, G., Diekmann, M., Dirnböck, T., Dörfler, I., Durak, T., De Frenne, P., ... Verheyen, K. (2018). Global environmental change effects on plant community composition trajectories depend upon management legacies. *Global Change Biology*, 24, 1722–1740. <https://doi.org/10.1111/gcb.14030>
- Peterka, T., Jiroušek, M., Hájek, M., & Jiménez-Alfaro, B. (2015). European Mire Vegetation Database: A gap-oriented database for European fens and bogs. *Phytocoenologia*, 45, 291–297. <https://doi.org/10.1127/phyto/2015/0054>
- Peyre, G., Balslev, H., Martí, D., Sklenář, P., Ramsay, P., Lozano, P., Cuello, N., Bussmann, R., Cabrera, O., & Font, X. (2015). VegPáramo, a flora and vegetation database for the Andean páramo. *Phytocoenologia*, 45, 195–201. <https://doi.org/10.1127/phyto/2015/0045>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19, 181–197. <https://doi.org/10.1890/07-2153.1>
- Pimm, S. L. (2021). What we need to know to prevent a mass extinction of plant species. *Plants, People, Planet*, 3, 7–15. <https://doi.org/10.1002/ppp3.10160>
- Prokhorov, V., Rogova, T., & Kozhevnikova, M. (2017). Vegetation database of Tatarstan. *Phytocoenologia*, 47, 309–313. <https://doi.org/10.1127/phyto/2017/0172>
- Purschke, O. (2017). *oliverpurschke/Taxonomic_Backbone: First release of the workflow to generate the taxonomic backbone for sPlot v. 2.1 and TRY v. 3.0*. Zenodo. <https://doi.org/10.5281/zenodo.845445>
- Reich, P. B. (2014). The world-wide 'fast-slow' plant economics spectrum: A traits manifesto. *Journal of Ecology*, 102, 275–301. <https://doi.org/10.1111/1365-2745.12211>
- Revermann, R., Gomes, A. L., Gonçalves, F. M., Wallenfang, J., Hoche, T., Jürgens, N., & Finckh, M. (2016). Vegetation database of the Okavango Basin. *Phytocoenologia*, 46, 103–104. <https://doi.org/10.1127/phyto/2016/0103>
- Ricklefs, R. E. (2008). *The economy of nature*. Macmillan.
- Rüsiņa, S. (2012). Semi-natural grassland vegetation database of Latvia. *Biodiversity & Ecology*, 4, 409. <https://doi.org/10.7809/b-e.00197>
- R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Samimi, C. (2003). Das Weidepotential im Gutu-Distrikt (Zimbabwe)—Möglichkeiten und Grenzen der Modellierung unter Verwendung von Landsat TM-5. *Karlsruher Schriften zur Geographie und Geoökologie*, 19, 1–136.
- Schaminée, J. H. J., Janssen, J. A. M., Haveman, R., Hennekens, S. M., Heuvelink, G. B. M., Huiskes, H. P. J., & Weeda, E. J. (2006). *Schatten voor de natuur. Achtergronden, inventaris en toepassingen van de Landelijke Vegetatie Databank*. KNNV Uitgeverij.
- Schmidt, M., Janßen, T., Dressler, S., Hahn, K., Hien, M., Konaté, S., Lykke, A. M., Mahamane, A., Sambou, B., Sinsin, B., Thiombiano, A., Wittig, R., & Zizka, G. (2012). The West African vegetation database. *Biodiversity and Ecology*, 4, 105–110. <https://doi.org/10.7809/b-e.00065>
- Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A., Reichstein, M., Bönisch, G., Díaz, S., Dickie, J., Gillison, A., Karpatne, A., Lavorel, S., Leadley, P., Wirth, C. B., Wright, I. J., Wright, S. J., & Reich, P. B. (2015). BHPMF—A hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography*, 24, 1510–1521. <https://doi.org/10.1111/geb.12335>
- Schultz, J. (2005). *The ecozones of the world*. Springer.
- Šibík, J. (2012). Slovak vegetation database. *Biodiversity & Ecology*, 4, 429. <https://doi.org/10.7809/b-e.00216>
- Sieg, B., Drees, B., & Daniëls, F. J. A. (2006). Vegetation and altitudinal zonation in continental West Greenland. *Meddelelser Om Grønland Bioscience*, 57, 1–93.
- Šilc, U. (2012). Vegetation database of Slovenia. *Biodiversity & Ecology*, 4, 428. <https://doi.org/10.7809/b-e.00215>
- Stančić, Z. (2012). Phytosociological database of non-forest vegetation in Croatia. *Biodiversity & Ecology*, 4, 391. <https://doi.org/10.7809/b-e.00180>
- Staude, I. R., Waller, D. M., Bernhardt-Römermann, M., Bjorkman, A. D., Brunet, J., De Frenne, P., Hédli, R., Jandt, U., Lenoir, J., Málš, F., Verheyen, K., Wulf, M., Pereira, H. M., Vangansbeke, P., Ortmann-Ajkai, A., Pielech, R., Berki, I., Chudomelová, M., Decocq, G., ... Baeten, L. (2020). Replacements of small- by large-ranged species scale up to diversity loss in Europe's temperate forest biome. *Nature Ecology & Evolution*, 4, 802–808. <https://doi.org/10.1038/s41559-020-1176-8>
- Stebler, F., & Schröter, C. (1892). *Versuch einer Übersicht über die Wiesentypen der Schweiz*. Landwirtschaftl Jahrbuch der Schweiz.
- Steinbauer, M. J., Grytnes, J.-A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H., Rixen, C., Winkler, M., Bärby-Durchhalter, M., Barni, E., Bjorkman, A. D., Breiner, F. T., Burg, S., Czortek, P., Dawes, M. A., Delimat, A., Dullinger, S., Erschbamer, B., Felde, V. A., ... Wipf, S. (2018). Accelerated increase in plant species richness on mountain summits is linked to warming. *Nature*, 556, 231–234. <https://doi.org/10.1038/s41586-018-0005-6>
- Synes, N. W., & Osborne, P. E. (2011). Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography*, 20, 904–914. <https://doi.org/10.1111/j.1466-8238.2010.00635.x>
- Testolin, R., Attorre, F., Borchardt, P., Brand, R. F., Bruehlheide, H., Chytrý, M., De Sanctis, M., Dolezal, J., Finckh, M., Haider, S., Hemp, A., Jandt, U., Kessler, M., Korolyuk, A. Y., Lenoir, J., Makunina, N., Malanson, G. P., Montesinos-Tubée, D. B., Noroozi, J., ... Jiménez-Alfaro, B. (2021). Global patterns and drivers of alpine plant species richness. *Global Ecology and Biogeography*, 30(6), 1218–1231. <https://doi.org/10.1111/geb.13297>
- Testolin, R., Carmona, C. P., Attorre, F., Borchardt, P., Bruehlheide, H., Dolezal, J., Finckh, M., Haider, S., Hemp, A., Jandt, U., Korolyuk, A. Y., Lenoir, J., Makunina, N., Malanson, G. P., Mucina, L., Noroozi, J., Nowak, A., Peet, R. K., Peyre, G., ... Jiménez-Alfaro, B. (2021). Global functional variation in alpine vegetation. *Journal of Vegetation Science*, 32, e13000. <https://doi.org/10.1111/jvs.13000>
- Trabucco, A., & Zomer, R. J. (2010). Global soil water balance geospatial database. CGIAR Consortium for Spatial Information. Available from the CGIAR-CSI GeoPortal at: <https://cgiarcsi.community/>
- van der Sande, M. T., Bruehlheide, H., Dawson, W., Dengler, J., Essl, F., Field, R., Haider, S., van Kleunen, M., Kreft, H., Pagel, J., Pergl, J., Purschke, O., Pyšek, P., Weigelt, P., Winter, M., Attorre, F., Aubin, I., Bergmeier, E., Chytrý, M., ... Knight, T. M. (2020). Similar factors underlie tree abundance in forests in native and alien ranges. *Global Ecology and Biogeography*, 29, 281–294. <https://doi.org/10.1111/geb.13027>
- Vanselow, K. A. (2016). Eastern Pamirs—A vegetation-plot database for the high mountain pastures of the Pamir Plateau (Tajikistan). *Phytocoenologia*, 46, 105. <https://doi.org/10.1127/phyto/2016/0122>
- Vassilev, K., Pedashenko, H., Alexandrova, A., Tashev, A., Ganeva, A., Gavrilova, A., Gradevska, A., Assenov, A., Vitkova, A., Grigorov, B.,

- Gushev, C., Filipova, E., Aneva, I., Knollová, I., Nikolov, I., Georgiev, G., Gogushev, G., Tinchev, G., Pachedjieva, K., ... Vulchev, V. (2016). Balkan vegetation database: Historical background, current status and future perspectives. *Phytocoenologia*, 46, 89–95. <https://doi.org/10.1127/phyto/2016/0109>
- Vassilev, K., Ruprecht, E., Alexiu, V., Becker, T., Beldean, M., Biță-Nicolae, C., Csörgő, A. M., Dzhovanova, I., Filipova, E., Frink, J. P., Gafta, D., Georgieva, M., Germany, M. S., Goia, I., Gumus, M., Hennekens, S. M., Janišová, M., Knollová, I., Koleva, V., ... Dengler, J. (2018). The Romanian Grassland Database (RGD): Historical background, current status and future perspectives. *Phytocoenologia*, 48, 91–100. <https://doi.org/10.1127/phyto/2017/0229>
- Vassilev, K., Stevanović, Z. D., Cušterevska, R., Bergmeier, E., & Apostolova, I. (2012). Balkan dry grasslands database. *Biodiversity & Ecology*, 4, 330. <https://doi.org/10.7809/b-e.00123>
- Vibrans, A. C., Gasper, A. L. D., Moser, P., Oliveira, L. Z., Lingner, D. V., & Sevegnani, L. (2020). Insights from a large-scale inventory in the southern Brazilian Atlantic Forest. *Scientia Agricola*, 77, e20180036. <https://doi.org/10.1590/1678-992x-2018-0036>
- von Wehrden, H., Wesche, K., & Miehe, G. (2009). Plant communities of the southern Mongolian Gobi. *Phytocoenologia*, 39, 331–376. <https://doi.org/10.1127/0340-269X/2009/0039-0331>
- Wagner, V. (2009). Eurosiberian meadows at their southern edge: Patterns and phytogeography in the NW Tien Shan. *Journal of Vegetation Science*, 20, 199–208. <https://doi.org/10.1111/j.1654-1103.2009.01032.x>
- Wagner, V., Spribille, T., Abrahamczyk, S., & Bergmeier, E. (2014). Timberline meadows along a 1000 km transect in NW North America: Species diversity and community patterns. *Applied Vegetation Science*, 17, 129–141. <https://doi.org/10.1111/avsc.12045>
- Walker, D. A., Breen, A. L., Druckenmiller, L. A., Wirth, L. W., Fisher, W., Reynolds, M. K., Šibík, J., Walker, M. D., Hennekens, S., Boggs, K., Boucher, T., Buchhorn, M., Bültmann, H., Cooper, D. J., Daniëls, F. J. A., Davidson, S. J., Ebersole, J. J., Elmendorf, S. C., Epstein, H. E., ... Zona, D. (2016). The Alaska Arctic Vegetation Archive (AVA-AK). *Phytocoenologia*, 46, 221–229.
- Wana, D., & Beierkuhnlein, C. (2011). Responses of plant functional types to environmental gradients in the south-west Ethiopian highlands. *Journal of Tropical Ecology*, 27, 289–304.
- Wang, Y., Heberling, G., Görzen, E., Miehe, G., Seeber, E., & Wesche, K. (2017). Combined effects of livestock grazing and abiotic environment on vegetation and soils of grasslands across Tibet. *Applied Vegetation Science*, 20, 327–339.
- Weigand, A., Abrahamczyk, S., Aubin, I., Bită-Nicolae, C., Bruehlheide, H., I. Carvajal-Hernández, C., Cicuzza, D., Nascimento da Costa, L. E., Csiky, J., Dengler, J., Gasper, A. L. D., Guerin, G. R., Haider, S., Hernández-Rojas, A., Jandt, U., Reyes-Chávez, J., Karger, D. N., Khine, P. K., Kluge, J., ... Kessler, M. (2020). Global fern and lycophyte richness explained: How regional and local factors shape plot richness. *Journal of Biogeography*, 47, 59–71.
- Weigelt, P., König, C., & Kreft, H. (2020). GIFT—A global inventory of floras and traits for macroecology and biogeography. *Journal of Biogeography*, 47, 16–43.
- Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, 199, 213–227.
- Whitfield, T. J. S., Lasky, J. R., Damas, K., Sosanika, G., Molem, K., & Montgomery, R. A. (2014). Species richness, forest structure, and functional diversity during succession in the New Guinea lowlands. *Biotropica*, 46, 538–548.
- Whittaker, R. H. (1975). *Communities and ecosystems* (2nd ed.). Macmillan.
- Willner, W., Berg, C., & Heiselmayer, P. (2012). Austrian vegetation database. *Biodiversity & Ecology*, 4, 333.
- Wiser, S. K. (2016). Achievements and challenges in the integration, reuse and synthesis of vegetation plot data. *Journal of Vegetation Science*, 27, 868–879.
- Wiser, S. K., Bellingham, P. J., & Burrows, L. E. (2001). Managing biodiversity information: Development of New Zealand's National Vegetation Survey databank. *New Zealand Journal of Ecology*, 25, 1–17.
- Wohlgemuth, T. (2012). Swiss forest vegetation database. *Biodiversity & Ecology*, 4, 340.
- WWF (2020). *Living Planet Report 2020-Bending the curve of biodiversity loss*. R. Almond, M. Grooten, & T. Peterson (Eds.). World Wildlife Fund.

BIOSKETCH

sPlot is a collaborative initiative to integrate existing local and national vegetation-plot datasets into a global harmonized database. It was initiated in 2013, within the sDiv working group 'Plant trait-environment relationships across the world's biomes'. Since then, it has become established as the largest vegetation-plot database worldwide and coordinates a consortium of 251 individual active members, representing 167 local and national datasets. sPlot's overarching scientific goal is the exploration of all aspects of global plant community diversity, including taxonomic, functional and phylogenetic diversity, across biomes, vegetation types, taxonomic or functional guilds and scales. Central to sPlot's mission is the exploration of the relationships between environmental drivers, trait variation, and assembly processes in local plant communities worldwide.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Sabatini FM, Lenoir J, Hattab T, et al. sPlotOpen – An environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecol Biogeogr*. 2021;00:1–25. <https://doi.org/10.1111/geb.13346>